DOCUMENT RESUME

ED 121 843 TM 005 275

AUTHOR Spitze, Hazel Taylor; Griggs, Mildred Barnes
TITLE Choosing Evaluation Techniques.
INSTITUTION Home Economics Education Association, Washington, D.C.
PUB DATE 27 Jan 76
NOTE 55p.
AVAILABLE FROM Home Economics Education Association, 1201 16th Street, N.W., Washington, D.C. 20036 (Stock No. 261-08424, $2.50)

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS Elementary Secondary Education; *Evaluation Methods; Grading; *Measurement Techniques; *Student Evaluation; *Student Testing; *Teaching Guides; Test Construction

ABSTRACT

This publication offers suggestions to any teacher in the field of education. It is not intended to answer all questions regarding evaluation, but rather to serve as a springboard toward stimulating new ideas and encouraging teachers to experiment with new methods of evaluation. Three classifications are used to show the varied methods of evaluation available: (1) real life situations, (2) simulations of reality, and (3) abstractions from reality. Chapter one details the process of evaluation, objectives, purposes, principles to guide the choice of techniques, unintended consequences, factors affecting the effectiveness of evaluation techniques, and the evaluation of cognitive, affective, and psychomotor domains. Chapter two discusses evaluation in real life situations such as observational techniques, rating scales, sociograms, checklists, interest and personality inventories, attitude scales, projective techniques, and interviews. Chapter three discusses simulation techniques in evaluation such as role playing, educational games, evaluation by inquiry, demonstrations, contrived incident, and Q sort. Chapter four discusses the construction of written tests, performance tests, oral tests, and other written exercises. Chapter five focuses on grades and their values. (DEP)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED121843

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICRO FICHE ONLY HAS BEEN GRANTED BY

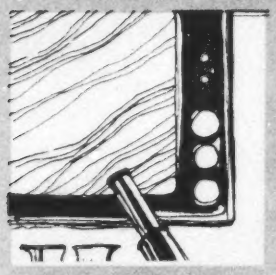
NEA

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM



Choosing Evaluation Techniques

Hazel Taylor Spitze
Mildred Barnes Griggs
Home Economics Education Association

TM005 275

Home Economics Education Association
1201 Sixteenth Street, N.W.
Washington, D.C. 20036

Executive Secretary
Catherine A. Leisher

Executive Committee
Leora N. Horning, President
Enid A. Carter, Immediate Past President
Lilla O. Halchin, Vice President
Stephanie H. Cooney, Secretary-Treasurer

Publications Committee
Mary Lee Hurt, Chairperson
Josephine Ruud
Reba J. Davis
Bonnie Paschold
Gwendolyn McQueen

Publication #261-08424
Additional copies are available.
For information write to Home Economics Education Association at above address.

Copyright © 1976 by Hazel Taylor Spitze and Mildred Barnes Griggs and the Home Economics Education Association. All rights reserved. Written permission required to duplicate any portion.

Library of Congress Card Catalog Number 75-32848

Cover design by Joyce W. Parkhill
Free-Lance Artist, Champaign, Illinois
Illustrations by Scott Griggs, age 12

Choosing Evaluation Techniques

by

Hazel Taylor Spitze
and
Mildred Barnes Griggs

Division of Home Economics Education
University of Illinois at Urbana-Champaign

HOME ECONOMICS EDUCATION ASSOCIATION
1201 Sixteenth Street, N.W.
Washington, D.C. 20036

Copyright 1976

A Message from the President of HEEA

The officers of the Home Economics Education Association are proud to provide you with another quality publication. This one was conceived, as are so many others, when someone said, "Wouldn't it be great to have a book of ideas relating to evaluation?" Without a doubt, Choosing Evaluation Techniques is just that! It adds to the long tradition of our organization for providing practical materials for teachers.

This publication offers helpful suggestions for teachers in any field of study. For this reason, HEEA is making it available to interested educators through our national headquarters office (address inside the front cover).

Choosing Evaluation Techniques is not intended to be a complete reference regarding evaluation, but rather a "spring board" of ideas to assist teachers in varying their techniques, encouraging their creativity and enlarging their evaluation perceptiveness.

HEEA is indebted to the two authors of this publication who so generously have given of their time and talent -- Hazel Taylor Spitze and Mildred Barnes Griggs of the University of Illinois at Urbana - Champaign, Department of Vocational and Technical Education, Division of Home Economics Education. Both have served as chairperson for the Division of Home Economics Education. HEEA is most grateful for their contribution.

We hope all readers will find this publication useful. The authors and our organization welcome your comments concerning Choosing Evaluation Techniques.

Leora M. Horning

Leora M. Horning, President
Home Economics Education Association

January 27, 1976

Table of Contents

	Page
A MESSAGE FROM THE PRESIDENT OF HEEA	2
ACKNOWLEDGMENTS	4
INTRODUCTION	5
CHAPTER 1—CHOOSING EVALUATION TECHNIQUES	7
The Process of Evaluation	7
Objectives	7
The Purposes of Evaluation	8
Clarification of Terms	9
Some Principles to Guide the Choice of Evaluation Techniques	9
Some Factors Affecting the Effectiveness of Evaluation Techniques	10
Unintended Consequences	12
The "Domains"—Cognitive, Affective, Psychomotor	12
Some Questions Teachers Often Ask About Evaluation	14
CHAPTER 2—EVALUATING IN REAL LIFE SITUATIONS	18
Observational Techniques	18
Anecdotal Records	18
Rating Scales	19
Peer Appraisal and Self-Report Techniques	21
"Guess Who" Technique	21
Sociogram	21
Interviews	23
Checklists	23
Personality Inventory	24
Interest Inventory	25
Projective Technique	25
Attitude Scale	25
CHAPTER 3—SIMULATION TECHNIQUES IN EVALUATION	27
Role Playing	27
Skits and Pantomime	29
Educational Games	30
Evaluating by Inquiry	30
Demonstrations	31
Laboratory and Experimentation	32
Contrived Incident	33
In-Basket	35
Discussion	36
Q-Sort	38
Writing Letters and Filling Out Forms	39
Audio-Visual Materials	39

CHAPTER 4—EVALUATING WITH TESTS 40
Written Tests 40
Constructing Written Tests 40
Table of Specifications 41
Sampling 42
Supply-Type Objective Items 42
Selection-Type Objective Items 43
Multiple Choice Test Items 44
Matching Test Items 45
Preparing and Scoring Essay Test Items 45
Performance Tests 47
Oral Tests 47
Other Written Exercises 47
CHAPTER 5—"PUTTING IT ALL TOGETHER" 48
Determining Grades 49
What Data to Include 49
Grades as Evidence of Growth or Achievement 51
Standards for Letter Grades 52
Relative Value of Scores 52
Concluding Comment 52

Acknowledgments

We are grateful to the following educators who kindly read the manuscript and offered valuable suggestions although we must assume the final responsibility for the contents: Mary Jo Clapp, Leora N. Horning, Bonnie Paschold, Gwendolyn McQueen, Dr. Ruth Hughes, Dr. Alyce Fanslow, Dr. Margaret Barkley, Dr. Mary Lee Hurt, Dr. Reba Davis; and to some of our present graduate students: Royston Lawson, Donna Woolcott, and Virginia Nash.

Introduction

This bulletin was written at the request of the Executive and Publications Committees of the Home Economics Education Association to be a "companion volume" to its 1970 publication, *Choosing Techniques for Teaching and Learning*. As the earlier volume stressed the need for variety in teaching techniques and some principles to guide our choices in that area, this one attempts to show that techniques for evaluating students are also widely varied and that a combination of techniques is more likely to result in a valid and fair appraisal than is any single one used alone.

Also like the earlier volume, this one is not written just for home economics teachers. The concrete examples illustrating the techniques will often be drawn from home economics, but since it is a broad field based in many disciplines, even the examples may be relevant for teachers in other fields. And certainly the techniques and principles will be applicable for all educators.

The classification scheme employed in *Choosing Techniques for Teaching and Learning* will be used again. Based upon the principle of reality, evaluation techniques will be classified as (1) real life situations, (2) simulations of reality, or (3) abstractions from reality, although it must be recognized that the way in which a technique is utilized affects its degree of reality.

We may think of the educative process as having four aspects: (1) objectives, (2) content needed to meet the objectives, (3) techniques for "getting across" the content needed to meet the objectives, and (4) evaluation to find out whether the objectives have been met. This publication will focus on the fourth aspect, evaluation, but of necessity will be concerned with the others as they relate to evaluation.

Our purpose will not be to provide answers to all questions raised but rather to stimulate thinking on the part of teachers and to encourage experimentation with new ideas and techniques. Instead of a recipe book, we offer some of our thinking and valuing and some of our experiences in evaluation. If, as a result of this bulletin, teachers are more critical of their tests, use more varied techniques in addition to tests, and do more thinking about the whole process of evaluation, our efforts will be rewarded.

HTS and MBG

Chapter 1

Choosing Evaluation Techniques

TEACHERS ARE OFTEN HEARD TO COMMENT, "I love to teach, but I don't like to evaluate." Actually, they probably would not "love to teach" if they had not found through some means of evaluation that their students were learning and hence that their teaching was successful. What they don't like is usually the assigning of grades. This publication is intended to help remove some of the frustrations while teachers wait for that millenium when grading is no longer required and to stress that the assigning of grades is probably the least important of the many reasons for evaluating.

The Process of Evaluation

What is evaluation? Evaluation is a cooperative, continuous *process* of making judgments, in terms of recognized objectives or standards and on the basis of appropriate recorded evidence, concerning past and present experiences (e.g., activities, situations, learnings, materials, processes), in order to affect the direction of future experience. As noted in the Introduction, it is the fourth aspect of the process of education.

Objectives

Without taking an active role in the current controversy regarding behavioral objectives, we would make these observations:

1. If you don't know where you're going, it's hard to get there!
2. When learning activities, evaluation, and reporting are all based on the same clearly defined objectives, the various components of the educative process are more likely to be consistent and to reinforce each other.
3. Teachers and students usually find it easier to assess progress when the objectives are clear, specific, and stated in terms of expected student outcomes, *BUT*
4. teachers tend to omit very important objectives that are difficult to measure when there is a great deal of stress on behavioral objectives which may often be trivial. And too many pre-specified objectives can eliminate or reduce flexibility and individualization in the classroom or other learning situation.

Objectives stated as learning outcomes provide direction for the teacher and the learner and serve as referents for evaluation. They can be stated in a variety of ways and can provide for cognitive, affective, and psychomotor outcomes at all levels, from the simple to the complex.

Gronlund[1] identified four ways in which objectives are often written that can cause the expected student learning outcomes to be *unclear*:

1. objectives that describe *teacher* behavior rather than *student* behavior,
2. objectives that simply list the subject matter to be covered,
3. objectives stated as learning processes rather than as learning products, and
4. objectives that are a combination of two or more outcomes.

[1]Norman E. Gronlund, *Measurement and Evaluation in Teaching*, 2nd ed. (New York: The Macmillan Co., 1971), pp. 7-8.

Objectives are most *meaningful* when they are:

1. stated in terms of desired student learning outcomes,
2. consistent with the fundamental structure of the subject matter,
3. realistic in terms of the abilities of the students,
4. in agreement with learning principles, and
5. representative of the needs of the learners.

The Purposes of Evaluation

The following may all be purposes of evaluation, and the priorities in a given situation vary according to the needs at that time:

- diagnosing learning strengths and weaknesses
- contributing to pupil learning
- improving instruction
- selecting teaching materials
- making judgments about appropriateness of objectives
- reporting pupil progress to parents
- using in guidance and counseling
- building self-concept (of students or teacher)
- using in school administration
- informing the public about the school
- seeking funding for special programs
- determining merit pay increases for teachers

Some of the foregoing may overlap depending upon how each is interpreted. In every case the key is in acquiring feedback. The concept of feedback is so central to evaluation that it was part of the title of an ASCD Yearbook which is one of the better resources for teachers on evaluation.[2]

What shall we evaluate as we utilize this evaluation process? Student learning is the first thing that comes to mind, of course, but we can and should also evaluate courses, programs, human relationships, teaching techniques, curriculum materials, and even the evaluation procedures themselves.

Questions that we might ask ourselves in thinking about criteria for evaluating our evaluative procedures include:[3]

1. Does our evaluation help us improve our teaching?
2. Are we clear on our purposes for evaluating?
3. Is the grading aspect placed in proper perspective?
4. What are we evaluating?
5. What means are we using to evaluate?
6. Are our objectives clear and used in evaluating?
7. By what standards are we evaluating?
8. Are we considering the uses to which the knowledge is to be put?
9. Are we using all the alternatives open to us in making our evaluations?

---

[2]Fred T. Wilhelms, ed., *Evaluation as Feedback and Guide* (Washington, D.C.: Association for Supervision and Curriculum Development, NEA, 1967).

[3]Adapted from a speech "Criteria for Evaluating Our Evaluative Procedures" by John Easley, Professor of Elementary Education, University of Illinois, to an extramural class in home economics education, 1968.

What shall we accept as evidence of growth in all the realms in which we evaluate? Borrowing a term, but not the technical definition, from the law profession, we might ask what is admissible evidence? Criteria for judging admissibility might include:

1. relevance
2. fact, truth
3. manner in which obtained (e.g., no invasion of privacy)
4. effect on one being evaluated
5. purpose for which used
6. source
7. form in which presented
8. fairness

Clarification of Terms

"Have you done your evaluations?"
"No, the exam is this week and grades have to be in next Friday."

In this two-sentence exchange, what can we conclude about the second speaker's definition or beliefs about evaluation? It appears that s/he uses the terms evaluation, testing, and grading synonymously. Other terms, like measurement, appraisal, and assessment, or formative and summative evaluation may also be confusing unless those who are discussing them come to some agreement about their meanings.

In this volume we shall try to avoid misunderstanding by using the following definitions:

Evaluation has been defined on page 7, paragraph 2. In another format we might define this process by answering the journalistic "w's" about it.

Who? instructor and learners and sometimes others
When? continuously or frequently
How? on the basis of evidence and in terms of objectives or standards
What? student learnings, teacher performance, curriculum, etc., etc.
Why? to affect future experiences

Measurement refers to the procurement of quantitative data.

Testing is the administration of scorable instruments which require the one being tested to write, speak, or perform a skill (or demonstrate a behavior).

Grading is the summarizing of a series of judgments to an abbreviation, usually a letter or a number.

Appraisal and **assessment**, if used, will have the same meaning as evaluation.

Formative evaluation is that which is done to improve the instructional methods and materials so that greater student learning will result.[4]

Summative evaluation is that which is done to appraise the overall effectiveness of a curriculum program.[4]

Some Principles to Guide the Choice of Evaluation Techniques

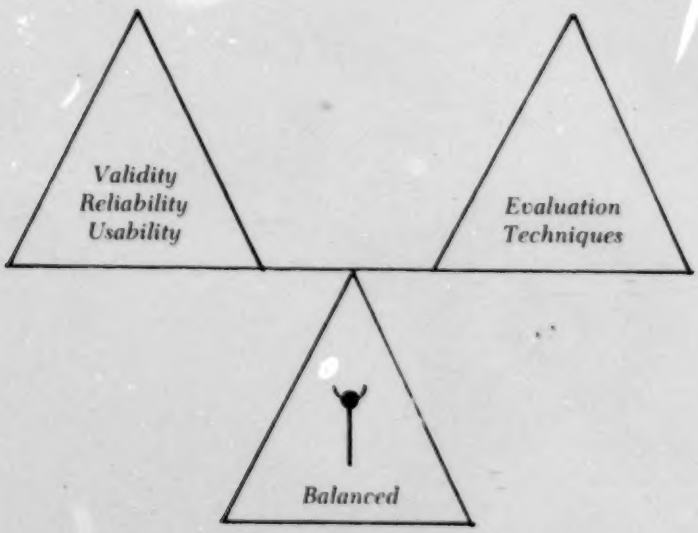
The following does not pretend to be an exhaustive list but it may provide helpful criteria as teachers make judgments about techniques to use in evaluation. We are using the term *principle* to mean a general factual statement of relationship. The order is random.

---

[4]Gronlund, *op. cit.*, p. 12.

1. The purposes for which evaluation results will be used affect the choice of evaluation procedures.
2. The greater the variety of evaluation techniques used, the more likely the judgment is to be a fair one.
3. The type of evaluation procedures used affects motivation and achievement.
4. The objectives for a course or other learning activity affect the procedures appropriate for evaluating that course or activity.
5. If the learner participates in choosing evaluation techniques, attitudes toward evaluation are more likely to be positive.
6. If evaluation procedures provide the learner with knowledge of objectives and evaluation results, his/her motivation and achievement are likely to be enhanced.
7. Unless a procedure measures the content it is said to measure, no judgment based on the measurement can be sound or fair.
8. If all who score a test arrive at the same score, judgments based on the test are more likely to be sound than when there are differing scores by different judges.
9. Clarity of directions can affect the validity of evaluation results with any given instrument.
10. The type of subject matter being taught affects the choice of evaluation techniques.
11. A learner's self-concept is affected by the evaluation procedures used and by his/her success or failure experiences with them.
12. When learners can utilize self-evaluation techniques, they are more able to identify their own needs and to take initiative in learning.

SOME FACTORS AFFECTING THE EFFECTIVENESS OF EVALUATION TECHNIQUES



Validity

In Principle 7 above we referred to evaluation procedures that "measure the content they are said to measure." In educational jargon this is called validity. There are several kinds of validity, and a teacher may be more concerned, at any given time, with one kind than another. If s/he wishes to see how much content students have learned, s/he wants results with *content* validity. If s/he wants to predict future behavior of the students in some area, s/he is interested

in *predictive* validity. If s/he wishes to have a measurement which concurs with that provided by another instrument, s/he looks for *concurrent* validity. Some now use the term *criterion-related* validity for both predictive and concurrent validity. If s/he wants to evaluate some psychological construct, such as reasoning ability, s/he checks for *construct* validity.

Whether one obtains valid results in an evaluation endeavor depends not only upon the instrument but on how, when, where, why, and with whom it is used. Hence, we cannot refer to a valid *instrument* (e.g., test) but only to valid *results*; validity exists only in relation to some specific use for the given results. Further, it exists on a continuum, not dichotomously as valid or invalid. For example, the results of a given evaluation with a given instrument may be highly valid for some students for some purpose but only moderately valid for some other purpose or some other students and of almost no validity for still other purposes or students.

Teachers in elementary and secondary schools can ask themselves frequently as they evaluate: Is this a valid measure of the behavior I expected when I wrote that objective? What evidence can this instrument (or activity) provide that this objective is being reached?

Such questions can help one to eliminate irrelevant test items and to add others, or to realize that a written test cannot obtain valid results in some situations. For example, if the educational objective refers to the ability to relate to young children, observing the student in interaction with young children is likely to provide more evidence than a multiple choice test. But observing beautiful relationships does not guarantee that the student understands the concepts and principles involved in teaching or parenting young children. Hence, there is the need for a variety of means of evaluation geared to the particular objectives.

Reliability[5]

Another factor affecting the effectiveness of evaluation techniques is reliability, or the consistency, of the measurement. If several persons scored the results, would all of them assign the same scores to each student? If so, the instrument or test is said to have high inter-rater reliability. (See Principle 8, page 10.) The less subjective the scoring is, the more likely it is that such consistency will be possible.

Written records are more likely to be reliable than anecdotes and impressions the teacher holds in his/her "memory bank." Guides for recording observations can aid in making this type of evaluation more reliable. Examples will be given in Chapter 3.

Reliability may also be affected by the length of the instrument used. Adding items may be the most feasible means of increasing reliability in some situations. On the other hand, deleting items that are unclear can also help.

Item Quality

In the case of a test, the quality of the items is of extreme importance. Differences in difficulty, all within a range that is reasonable for the student being tested, and appropriateness for the outcome being measured, help to improve test quality. Unclear, ambiguous items or items which "give away" the answer cannot result in valid measurement. Likewise, a too-rigid key used in scoring can produce problems. For example, the clothing teacher who included a fill-in-the-blanks item, "Clipping is _____ _____ _____ _____," and scored with a key that read "necessary to prevent puckering," might be challenged by a student who answered "a penalty in football"! Either the key, or, preferably, the stem of the item needed revision.

---

[5]For an in-depth discussion of the reliability of measurement methods, see J. Stanley Ahmann and Marvin D. Glock, *Evaluating Pupil Growth: Principles of Tests and Measurements*, 5th ed. (Boston: Allyn and Bacon Inc., 1975), ch. 10.

Usability and Administration

Teacher-student relationships, including non-verbal communication between them, can be a factor in evaluation results, especially the teacher's expectations for the student.

The physical arrangement of the furniture, the location of students in relation to each other, the temperature, humidity, ventilation, lighting, and other factors relating to comfort—all can affect evaluation results.

The time available to carry out the evaluation procedure, including both that for the student to demonstrate his ability and that for the teacher to score, can affect results.

The anxiety level of the student and the state of his/her health at that particular time can have important effects. High levels of anxiety can often be reduced by the manner in which the procedure is carried out and by not placing too much importance on any single activity. Extremely high levels of anxiety and the complete absence of anxiety can both have negative effects for the student.

Physical aspects of printed evaluative instruments which have effects include: (a) length and appearance, (b) clarity of directions, (c) clarity of print (Is *every* copy readable?), and (d) reading level. Unless one's objective is to measure vocabulary and reading level rather than understanding of concepts and principles, tests and other evaluative instruments should be written on a reading level that the slowest student can understand. For many high school students and adults, this may mean 4-5 grade level.

Unintended Consequences

Sometimes there are unintended consequences of evaluation efforts, too. Motivation may be increased or decreased, and attitudes toward the subject, toward the teacher or school, or toward learning in general may be affected.

There may also be important effects on the self-concepts of the students. A "Peanuts" cartoon[6] a few years ago illustrated this well. Linus explained excitedly to Charlie Brown how the home team had won the game in the last three seconds of play and how fantastic the players felt, and Charlie Brown unexpectedly reminded him of the *other* team's feelings. Replayed in a school setting the sequence might read:

> Charlie Brown, I just had the most wonderful experience. We had this pop quiz. . .
>
> You think *that's* wonderful??
>
> Yes, absolutely wonderful! I made the HIGHEST in the class. You wouldn't believe how lucky I was! I guessed nearly every question right. And the teacher had me stand up and she told everyone I had the highest and I made A+. It was fantastic!
>
> How did the *other* guys feel?

The "Domains"—Cognitive, Affective, Psychomotor

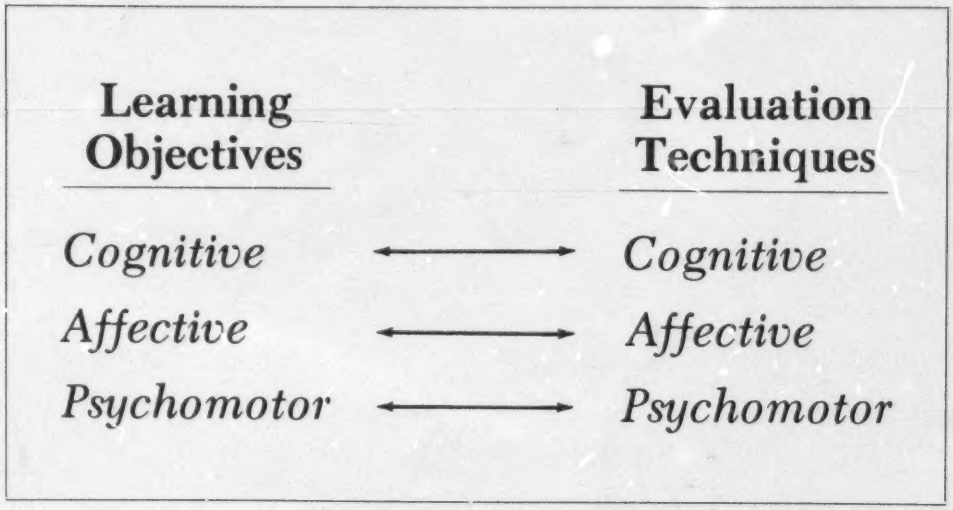
Teachers are well aware that not all educational objectives have to do with the acquisition of information or thinking and reasoning. These are only the cognitive ones.[7] And we all know, too, that cognitive objectives range over several levels from mere repetition of facts, through comprehension and application of knowledge, to analysis, synthesis, and finally, evaluation.

It is usually much easier to find out whether students have memorized some definitions and lists of categories than it is to determine whether they can see the relationships and make

---

[6] Charles Schulz, "Peanuts featuring 'Good ol' Charlie Brown" in Champaign-Urbana (Illinois) *Courier*, Oct. 26, 1969.

[7] For a complete description and analysis of this domain see Benjamin Bloom, ed., *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain* (New York: David McKay Company, Inc., 1956).



any use of the knowledge. Teachers have tended to have a preponderance of test items at the lowest cognitive levels. If we are careful to include objectives at all six cognitive levels, and if we keep these objectives in front of us as we evaluate, perhaps we can avoid this pitfall. In the chapters that follow, we shall endeavor to indicate how the evaluation techniques to be described can be used at various cognitive levels.

Affective objectives,[8] too, must be a part of our curriculum planning and hence our evaluations. After the strong emphasis in the late fifties and early sixties on cognition, we had to rediscover that attitudes, interests, appreciations, feelings, and values are important in education. We even had to relearn that certain feeling, or affective states, make the acquisition of knowledge and the development of thinking ability difficult if not impossible.

Insisting upon affective objectives in the curriculum is not to suggest that educators should determine for students what they *ought* to value, appreciate, or be interested in. We do feel that teachers have a responsibility for helping students to be aware of their own values, to understand the sources of these values, and to predict the consequences of their values, as well as to see the need for tolerance of the values of others.

Evaluating in this domain is difficult! Sometimes evidence concerning some affective objective appears quite unexpectedly, and we need a procedure for recording it, perhaps some type of anecdotal record. It may be that indirect observations, rather than direct attempts at measurement, are more valid in these areas. If you want to know how a person feels about something, is it better to ask him/her or to observe behavior when s/he doesn't know you're thinking about those feelings? We don't think we know the answer to that question, but the old saying, "your actions speak so loudly I can't hear what you say," is not without merit.

In the descriptions of evaluation techniques in Chapters 2-4, we shall try to point out ways in which some of them can be used in the affective domain.

Teachers have usually been more aware of the need to evaluate in the psychomotor area[9] than in the affective. We knew that we were supposed to teach skills, and we felt that we should measure, somehow, the level of skill that our students attained. Often, though, we have

[8]See David R. Krathwohl, Benjamin Bloom, and Bertram B. Masia, *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook II: Affective Domain* (New York: David McKay Company, Inc., 1956).

[9]One source in this area is Elizabeth Jane Simpson, "The Classification of Educational Objectives, Psychomotor Domain," a project supported by a grant from the U.S. Office of Education in which Mildred Griggs, Nancy Carlson, and Carol Hodgson served as assistants. It was published in *Illinois Teacher of Home Economics*, Vol. X, No. 4 (1966/67).

Another source is Anita J. Harrow, *A Taxonomy of the Psychomotor Domain, A Guide for Developing Behavioral Objectives* (New York: David McKay Co., Inc., 1972).

tended to equate level of skill with the quality of a product just completed, and we have stressed product rather than process.

If a student in an adult class "turns out" a beautifully tailored suit, can we assume that s/he has the skills to do another one? We thought so until we observed the same students enrolling every fall in the same tailoring class and we made some inquiries. We found students dependent upon the teacher (who seemed to enjoy the dependency) to check every step, and in some of the more difficult steps, like putting in a smooth sleeve, the teacher was actually doing the work! She was "demonstrating" on the student's garment, over and over. So the product was faultless, but the student had not learned the process.

The reverse could also be true. If a student lays a pattern off-grain and learns that his/her garment does not "hang right" nor fit very well, can we assume that s/he has not acquired the skill of placing a pattern on grain? We think not. One can learn from one's mistakes—if they are not fatal!

In both of the above examples, values are also involved. In the first case, both teacher and student placed a high value on a "perfect" garment and a lower value on the learning. In the second, both valued the learning which could be applied to future garments and were less concerned about the garment on which they were practicing to acquire these skills.

Likewise, the cognitive area is a part of both situations. Knowledge is needed at several levels in order to be successful in these endeavors. Hence, we can see the inter-relationships among the "domains." Seldom do we have a learning situation that involves only one, and often we have situations which involve all three domains. We may be able to evaluate all three in one operation, but it is likely that the situation may require several procedures to obtain an assessment in cognitive, affective, and psychomotor domains.

Some Questions Teachers Often Ask About Evaluation

1. What is the difference between evaluating in terms of objectives and evaluating by comparing with a standard?

If the objective is stated in such a way as to incorporate the standard, there may be no difference, and sometimes this is appropriate. If a certain test for employment which members of the typing class are preparing to take requires a speed of sixty words per minute with no errors, then the standard is set for them outside the class. The objective may be to pass the test, and anything less than the standard may be failure. On the other hand, each student may have an objective of developing the greatest speed and accuracy of which s/he is capable during the time allowed in the course. In this case the student who reaches sixty words per minute with no errors will not relax and feel "finished" nor will the student who reaches forty-five words per minute with two errors feel an utter failure. Instead, s/he might feel simply that a few more weeks are needed to prepare for the test.

Let's take another example. In a food preparation class, students may be told that one of the objectives is to develop the skill to make standard biscuits. If they are preparing for employment in a bakery where "standard" biscuits are sold, they need the knowledge of the size, shape, texture, color, and flavor of standard biscuits, and they need the skill to make them. But if they are preparing to cook only for themselves or their own families, the "standard" biscuit may not be their special need. The standard was developed by those who liked biscuits two inches high, light golden brown, etc., but their family may prefer biscuits one inch high, dark, and crusty. They may need to know how to vary the standard recipe and instructions in order to please those for whom they are cooking. If we tell them, even nonverbally, that only the standard biscuit is "good" and that people who like non-standard biscuits are a bit strange, we can do damage to self-concepts that may already be weak.

What about a standard in a non-skill area? If those who score 93-100% on the history or math test make A, those who score 85-92% make B, etc., are we using an outside standard to

evaluate? Whose standard is it? And if this outside standard is applied to every teacher-made test, does it make any sense?

It was probably in response to the objections implied in these questions that the "bell-shaped curve" came into use. If my test is not perfect, the teacher could say, then the scores may go up or down but no student will be punished because I "grade on the curve." But other questions then arise. Bell-shaped curves apply to average groups, or random samples of a total population. Is *this* group average? Should the top of the curve for this group be C or B?

We also get enmeshed in the question of speed or power. Is the student who can score 90 on the test in an hour not as good as the one who can score 90 in forty minutes? When is it appropriate to measure speed?

If standards are used in making evaluations, appropriate questions to ask might be: (1) Why is this standard used? (2) Whose standard is it? (3) What will the use of this standard do *to* or *for* the students?

Some standards are set up by the society through some agency of government for the protection of all (e.g., those necessary to pass a test for a driver's license), some by professional or trade associations for entrance into the group, some by a particular type of employer or a single employer, some by a single teacher in a classroom. Some may be set by the students themselves. Some are quite appropriate and some are questionable. Answering the questions in the preceding paragraph may help to avoid those that are arbitrary and a possible hindrance in the educative process.

2. Must all evaluations be included in the student's grade?

Evaluation is needed for many purposes (see page 8) and grading is only one of these, so the answer to the above question has to be No. We tend, however, to leave out those evaluations which are not "part of the grade" and to wonder a great deal about which ones should be a part of the grade. There are eminent educators[10] today who are proclaiming that there should be no grades. Their objections are based, partially at least, on the fact that a single letter cannot carry the meanings that evaluations should give to students, parents, and others. This abbreviation is highly confusing. What does a grade of C mean? It is almost like using "c" in speaking and leaving the listener to wonder whether it meant correct, careful, cloudy, clean, crimson, or something else, or even whether it was an adjective at all! Perhaps instead of trying to abbreviate everything into one letter we should, as a few schools have done in a variety of ways, list several aspects of performance and evaluate each with some descriptive term or comment. For example, rather than B in math, the student might see on his report card, with a comment after each:

Understanding of mathematical processes
Accuracy in computation
Ability to apply processes to everyday problems
Appreciation of the beauty of mathematics
Effort expended in learning mathematics
Interest in the history of mathematicians
Appreciation of the usefulness of mathematics
(or whatever was appropriate for a particular math course)

Even a short list could help a student, and the parents, understand more about his/her strengths and weaknesses and give some guidance on how to improve, if s/he wanted to improve.

"But," protests a teacher, "at this point in time we have to use the single letter abbreviations. How do we decide what to include in the 'B'?" This is a professional judgment, and a teacher is a professional person with professional responsibility. Just as an engineer has to decide what materials and what processes to use in spanning a river with a bridge, so a teacher

[10]For some examples, see chapter 5.

has to decide what materials and processes to use in evaluating students. The background on which s/he makes the decisions comes from the disciplines on which education is based, such as psychology, sociology, and philosophy.

Should a student's attitude in class affect the grade? It is easier to ask such questions than to answer them. It is our opinion that the effort a student puts forth (and this reflects attitude) should be a part of the grade if all the evaluation has to be expressed in a single letter. We think that a student's success in a job or an activity involving writing, for example, depends not only upon his/her knowledge of the rules of grammar but also upon enjoyment of writing, feelings about words, desire to communicate clearly and simply or eloquently. Hence, we see affective and cognitive objectives and evaluations inextricably interwoven.

Which is more important, knowledge or skill? Should the student who can "sew a fine seam" or the one who can score high on a written test make the higher grade in a clothing class? That depends upon the objectives which the teacher and students have set for themselves in the class and also on how the grades will be used. If a mentally retarded student in an occupational class learns how to sew on buttons and is able to handle a job at a dry cleaning shop where s/he sews on buttons all day, s/he should be considered successful. Another student who wishes to learn to sew as an expression of creativity would need a much wider range of skills and also some knowledge of fabrics, patterns, fit of garments, color, and the like. Still another whose desire to sew stemmed from a need to save money would require different knowledge and skills, especially if time were also at a premium.

3. How does a teacher differentiate between the A, B, and C?

No one can give a teacher a formula for such differentiation that will work in every instance. However, certain principles are relevant to this differentiation and, if applied, may be helpful. When the difference between the lowest B, for example, and the highest C is very narrow, the reliability of the letter grades tends to be much less than when the gaps are wider. For example, if the cut-off point for B is a score of 39 and the highest C had a score of 37, it will be much harder to defend the differentiation than when the cut-off for B is 39 and the highest C is 29. Also, the differentiation will be more defensible if many types of techniques are used and every student has many chances to show his/her growth and strengths. If great care is taken to weigh the different components of the final letter grade according to their relative importance in the total course or learning experience, it is less likely that questions will be raised about the grade. Student self-evaluations and peer evaluations may also contribute to a greater sense of fairness about the final grade. Having recorded evidence of various kinds may also facilitate the judgment that is the ultimate differentiation between letter grades.

4. Should evaluation be objective, that is, avoid subjectivity?

Dictionaries provide such definitions of evaluation as "the process of appraising carefully, or of ascertaining the value or amount of, or of estimating the value of." For our definition of educational evaluation, refer to page 7. Could we not say that, by definition, evaluation can*not* be objective? It is a judgment. In educational evaluation, we hope it is a professional judgment (although, of course, laypersons make judgments about educational matters, too). BUT, we do want our judgments to be based on as much objective evidence as possible.

What constitutes objective evidence? Is a recording made by a music or speech student, or a video tape of a dance routine or a ball game, objective evidence? Is a score on a test, or a diary record of baby sitting experience, objective evidence? What about a "term paper," a lamp base from the woodworking class, a salad or a loaf of bread, a painting or craft object, an oral report of a field trip, a book review, a record of income and expenditures? All these things could provide evidence for evaluations of one kind or another. Their objectivity is a matter of degree and depends partly on the situation. In every case, though, there are also elements of subjectivity. Someone has to judge the recording or the term paper or the art object. Someone made the test, and even the so-called objective test is subjective in the sense that the maker judged which items to include, how many items it should contain, how long the students should

have to take the test, and the like. In reality, only the scoring is objective, and for some tests, that, too, becomes subjective.

Nevertheless, we should strive for objective evidence and seek as wide a variety as possible within the constraints of the situation. And we should constantly try to strengthen our ability to make reasonable judgments just as we hope the physician continually improves his ability to diagnose our ills and prescribe remedies. Such competence can be increased by trial and error if we study carefully the situation in which we are trying and erring, but there are also other helps available in professional literature, conferences, and discussions with colleagues and consultants.

5. How do we evaluate a student who has done nothing?

The teacher who asks this question has already evaluated, i.e., made a judgment, in order to say that the student has "done nothing." If the question is really "What grade do we assign?," then the answer must surely be a failing one if the judgment of "done nothing" is a fair one. We need, of course, to note much earlier that the student is doing nothing and try to find out why, offer alternatives, and do everything we reasonably can to get him/her into a pattern of productivity. If a student does not learn, then the teacher has not taught, so a failing student does not signify a successful teacher, and even though we know we cannot always succeed with every student, we don't settle for less until we have tried all of our skills and techniques.

6. Who should do the evaluating?

Anyone who had a part in setting the objectives is already automatically involved in the total process of evaluation. As evidence is collected to ascertain whether the objectives are being met, both teacher and students are intimately involved. Sometimes others, e.g., parents or administrators or employers, may also play a part.

Within the classroom both teacher and students can participate in making the necessary judgments concerning the evidence collected, as well as in setting objectives and deciding what evidence to collect. Students may judge their own work and assist in judging that of their peers if the climate and interpersonal relationships are satisfactory. They may also participate in judgments about the curriculum and the teaching. We wonder what effect it would have on the morale of the students and on the teacher-student relationships if, on the day that report cards are given out to the students, they also gave the teacher a report card! In their own language they might offer judgments on:

Is prompt and always present
Knows the subject and teaches what is important
Is well prepared for class each day
Likes us
Is fair
Makes class interesting
Etc.

7. When should evaluation be done?

In a word, *often!* If not continuous, the evaluations can be frequent, and feedback to students can lead to new desires to learn.

8. How can pre-assessments help?

If students, through pre-tests and other means, can discover that which they do not know and develop a curiosity to gain this knowledge, the effect on motivation can be significant. Puzzles, games, and contests can sometimes lead to this discovery and development of curiosity, and certain kinds of quizzes and projective techniques can also be motivational in this regard.

Chapter 2

Evaluating in Real Life Situations

STUDENTS ARE GENERALLY EAGER TO PARTICIPATE in learning activities that are part of their real life situations because they can apply the content learned in the classroom to real events. Persons, such as employers, peers, parents, teachers and supervisors, who are affected by or involved with these real life situations can help students evaluate their performance.

Student performance in real life situations can generally be evaluated by one or more of the following procedures: (1) observing as a student plans, and prepares for an activity, (2) observing as a student performs an activity and describing and/or judging the behavior, (3) observing and/or judging the quality of the product resulting from the activity, (4) asking others who were involved in or affected by the activity about the performance, and (5) questioning the student directly. These procedures require direct observation of the learner and can be facilitated through the use of observational techniques, peer, supervisor, or teacher appraisals, and self-report techniques.

These techniques require the observer to make subjective judgments about the learner. However, we are faced with the choice of either using them to evaluate student performance and growth as validly as possible or ignoring those outcomes that cannot be measured by conventional testing methods.

Objectives stated as expected student outcomes can guide us in determining what behaviors to focus on during observations. The following questions may be useful in determining what behaviors to observe in real life situations.

1. Can this behavior be evaluated by other means?
2. Is there a conflict between teacher expectations and the culture (religious, ethnic, racial, regional) within which the student is interacting?
3. Can I make a fair interpretation of the student's behavior?
4. What standards shall I use to judge the student's performance?

Observational Techniques

Some frequently used observational techniques are anecdotal records and rating scales.

Anecdotal Records

Each student tends to have some unique characteristics that attract the teacher's attention. We are aware, for example, of the students who seem bored with school and life in general, the ones who need special attention and verbal praise, the ones who plunge right into their work at the sound of the bell, the ones who dawdle, the ones who appear very sullen and unapproachable, and the ones who are generally pleasant. We can usually identify those who do their own work and those who most often depend on others. We know the students who are often tardy and those who are punctual. We know our poor readers and slow learners just as we do those who grasp ideas readily. Such observations tell us how students typically perform and serve as a guide for us to measure desired behavior changes. Therefore, it is important to keep a record of

behaviors in order to assess changes. An anecdotal record is a way to keep an account of our observations and the impressions gained through the observations.

Anecdotal records are factual descriptions of meaningful behaviors and incidents of students observed and recorded by teachers. Gronlund[1] suggested that (1) records be made as soon after the observation as possible, (2) factual description of incident and observer's interpretation of incident be kept separate, (3) positive and negative incidents be recorded, and (4) inferences concerning typical behavior be drawn only after a number of observations have been made.

Anecdotal records of behaviors in real life events of students can be meaningful. One may wish to observe a student's behavior on the job, as a class member, as a person, etc. An anecdotal record entry may be as follows:

Class: *Child Development* Student: *John James*
Date: 4/30/75 Place: *C.D. Classroom* Observer: *MG*

Observation

When the children entered the play school lab, young Bobby began to cry and run after his mother. John gently lifted Bobby into his arms and began talking about the great time they were going to have at school that day. When Bobby stopped crying John involved him in free play with the other children. John smiled and appeared pleased with himself.

Interpretation

John seems to be much more confident of his ability to handle the children. A few weeks earlier he ignored a similar incident and informed me that he didn't know what to do.

If recording time is limited, a more abbreviated version of this incident could read:

Observation

John seemed please when he quieted Bobby toda and involved him in free play.

Interpretation

John's confidence and ability to handle the children has increased.

Rating Scales

A rating scale is a reporting device that has a list of characteristics or qualities to be judged and a scale for indicating the degree to which each attribute is present. The characteristics or qualities to be judged should be observable and consistent with the objectives. Gronlund[2] classified the uses of rating scales into three major evaluation areas: (1) procedure (process), (2) product, and (3) personal-social development.

In real life situations a rating scale may be used to evaluate student behavior as a consumer, club member, friend, committee member, family member, a volunteer, a paid worker, a tutor, etc. The characteristics or qualities to be judged can be adapted to either a numerical rating scale, a graphic rating scale or a descriptive-graphic rating scale.[3] It is most meaningful to share this data with students in a conference so ratings can be fully explained. Student

---

[1]Norman E. Gronlund. *Measurement and Evaluation in Teaching* (New York: The Macmillan Company, 1971), p. 416.
[2]*Ibid.*, p. 421.
[3]*Ibid.*, p. 418.

acceptance of rating, and their personal growth will be increased if they are informed of the criteria used to make the judgment.

A *numerical rating scale* requires the observer to use numbers to indicate the degree to which a characteristic or quality is present.

Example

Directions: Indicate the degree to which this student assumes responsibility as a teacher aide by circling the appropriate number. The numbers have the following values:

5—always, 4—frequently, 3—occasionally, 2—seldom, and 1—never.

1. To what extent does the student help the children with their art projects?

   1 2 3 4 5

2. To what extent is the student's help appropriate?

   1 2 3 4 5

The *graphic rating scale* has each characteristic or quality that is being rated on a horizontal line. The observer places a mark on the line to indicate the degree to which the characteristic is present.

Example

Directions: Indicate the degree to which this student assumes responsibility as a teacher aide by placing an "x" at the appropriate place on the horizontal line.

1. To what extent does the student help the chilren with their art projects?

never	seldom	occasionally	frequently	always

2. To what extent is the student's help appropriate?

never	seldom	occasionally	frequently	always

The *descriptive rating scale* contains brief statements in behaviorial terms along the horizontal line where the observer can rate student performance. Statements may be placed at each designated point or only at the center and end positions. Space is frequently left for comments by the rater.

Example

Directions: Indicate the degree to which this student assumes responsibility as a teacher aide by placing an "X" at the appropriate place on the horizontal line.

1. To what extent does the student help the children with their art projects?

never helps; quiet, passive	helps when asked	helps without ever being asked

2. To what extent is the student's help appropriate?

does all the work for the children without encouraging them	often does work for the children when they have difficulty	Provides guidance and direction. Let's children do work.

Comment:

Peer Appraisal and Self-Report Techniques

An important function of evaluation is providing feedback to students for their own personal growth and development. Another function is providing data for teachers to use in structuring the learning environment to promote optimum growth and development for each student. Who should make these evaluative judgments that have the potential to play such a crucial role in the lives of students? Hopefully the teacher is involved in all areas of evaluation, but often it is advantageous to have the teacher's evaluation supplemented by others, particularly students' peers, who observe and participate with them in real life situations. Students quite often know the strengths and weaknesses of each other from daily contacts. Their work and play relationships force them to make evaluative judgments regarding the way their peers get along with others, respect the rights of others, accept responsibility for their behavior and that of the group, assume leadership roles, and other behaviors.

The same evaluation techniques that teachers use to evaluate student behavior in real life situations can be used by their peers. However, two other techniques may be particularly successful for peer appraisal: (1) the "guess who" technique, and (2) the sociogram.[4]

"Guess Who" Technique

The technique is a simple procedure in which a list of brief positive behavior descriptions are handed to each student. The students are asked to name the class member who best fits each description. They may name as few or as many students as they wish for each item. Examples of items that could be used are:

1. Here is someone who is not argumentative but likes to talk a lot.
2. Here is someone who is kind and seems genuinely concerned about the welfare of others.
3. Here is someone who is friendly, has lots of friends and tries to make others feel comfortable.

The technique is scored by counting the number of times each student is mentioned on each description. This technique can indicate the reputation students hold among their peers and can make teachers aware of feelings among students that they had not detected.

Sociogram

The sociogram is a means for assessing the social structure of a group. It shows some measure of social acceptance because it is based on student's selection of companions for group activities. When student choices are made for activities in which successful completion is dependent on skill and knowledge, social acceptance does not tend to be a major factor in selection. Gronlund[5] made the following suggestions for teachers who wish to use this technique.

[4] *Ibid.*, p. 436.
[5] *Ibid.*, p. 441.

(1) The situations in which choices are made should be a normal part of the ongoing learning sequence. (2) The basis for choice (e.g., food preparation lab) and restrictions on the choosing (e.g., number to be chosen) should be made clear. (3) All students should be free to participate in the activity. (4) The choices made by the students should be kept confidential. (5) The choices made should be used to organize the groups unless it produces cliques or other undesirable consequences.

There are many opportunities for using the sociometric technique. Laboratory groups, field trip teams, group projects, and committees can all be organized in this way. Theoretically, groups should be expected to function more smoothly when students have been permitted to select those persons with whom they can work well. One limitation, however, may be that total classroom or group interactions will be restricted and students cannot make choices until they know each other well. It is not appropriate to use it on every occasion when grouping is desirable.

Example

Sociogram Form

Name_____ Date_____ Class_____ Period_____

During the next four weeks we will be engaging in activities that will permit you to work in a group with three other people each week. I think that you know everybody in this class well enough to select persons whom you want to work with in play school lab, visiting a nursery school, making toys for the lab and making an oral report. You may choose a different group for each activity. I will keep your choices confidential and assign you to participate with three of your four choices whenever possible.

I choose to work in the lab with:

1. _____ 2. _____ 3. _____ 4. _____

I choose to go to a nursery school with:

1. _____ 2. _____ 3. _____ 4. _____

I choose to make toys with:

1. _____ 2. _____ 3. _____ 4. _____

I choose to make an oral report with:

1. _____ 2. _____ 3. _____ 4. _____

Self-report techniques can provide information about students' past behavior and experiences and data that are not observable such as one's interests, beliefs, attitudes, feelings about self and others, opinions, and values. When these evaluation techniques are used one assumes that students are willing and able to provide accurate information. There is always the possibility that students' recollection of past events and perceptions of themselves will be biased, as is true for most people.

The information collected through the use of self-report techniques can help teachers better understand individual students and plan to meet their needs. It can be used to assess readiness for new experiences and to make curriculum changes.

Interviews

Interviewing students as a self-report technique can be very time consuming; however, the one-to-one relationship between teacher and student allows for a great deal of spontaneity and flexibility. Teachers can get clues about students' strengths and weaknesses by the way they respond to questions, ask questions, evade issues, or share information. If a trusting relationship has been established, students may be willing to communicate more information via an interview than they will put into writing. This may be an advantage for students who do not express themselves as well in writing as they do orally.

Checklists

A checklist is an observational technique that requires the observer to provide a "yes-no" judgment.

Checklists are similar in appearance and use to rating scales; however, rating scales require judgments about the degree to which a characteristic is present. Teachers should be careful not to use a checklist when they want evaluative data that is more indicative than "yes-no" or something equally as limited.

Checklists can consist of lists of characteristics that pertain to real life situations. The observer can indicate whether the student possesses the characteristic.

Example

Classroom Citizenship Checklist

Student _____ Class _____

Teacher _____ Date _____

Place an X in the blank preceding the statements that best describe the student named above.

____ 1. Is on time for class most of the time.
____ 2. Has necessary supplies for class most of the time.
____ 3. Is generally courteous to teacher and classmates.
____ 4. Accepts suggestions and help.
____ 5. Shows respect for property of others.
____ 6. Makes good use of time.
____ 7. Hands in assignments on the date due.
____ 8. Is attentive during most of class period.
____ 9. Avoids disturbing others when they are working.
____ 10. Is sensitive to the needs of others.

Activity Checklist

This technique can be very useful if a teacher is trying to determine information about students' earlier learning experiences. It is often useful to know which experiences are common among all students and which ones are unique to certain students. In a child development class, for example, a teacher could compile a list of activities related to caring for children and ask the students to indicate the things they have done. The results can be used for planning future learning experiences.

Example

Child Care Activity Checklist

Name_____ Date_____ Period_____

Directions: Indicate by placing an "X" in the blank opposite the things that you have done before.

____ 1. Diapered a baby	____ 4. Dressed a baby
____ 2. *Prepared a baby's formula*	____ 5. *Heated a baby's bottle*
____ 3. Bathed a baby	____ 6. Folded diapers

Problem Checklist

Data regarding the kinds of problems students encounter in their daily lives can be valuable for planning courses that deal with personal, social and emotional development. The checklists generally consist of a list of problems common to people of various age ranges, sex, social class, etc. Students are asked to indicate the particular problems which trouble them. The list might include items such as:

1. arguing a lot
2. having too few friends
3. having too little money
4. doing poorly in school work
5. eating too much

If it is necessary to have students identify themselves by names on the checklist, they should be free to refuse to participate.

Personality Inventory

A personality inventory consist of a series of questions which students respond to by indicating "yes," "no," or "uncertain," or "strongly agree," "agree," "undecided," "disagree," or "strongly disagree." An inventory generally consists of items that are often of concern to students.

Responses to the one illustrated below can indicate whether persons feel mainly positive or negative about themselves, their emotions, values, academic performance, special abilities, appearance, self-governing and achieving abilities, and their abilities to get along socially and establish satisfying relationships with parents and families.

Student responses to individual items or groups of items may be used as a basis for conferences.

Example

A Checklist About Me[6]

	strongly agree	agree	undecided	disagree	strongly disagree
1. I get upset often at home.					
2. I am an interesting person to other people.					
3. It takes me a long time to get used to anything new.					
4. I do very well in school.					
5. I don't like the way I look.					

[6]See Ruth Harms, "A Checklist About Me," *Illinois Teacher*, Vol. XVII, No. 2 (Nov.-Dec., 1973), 76-82.

Interest Inventory

Interest inventories have been used primarily for educational and vocational guidance. Students are generally asked to respond to items about school subjects, occupations, leisure time activities, people they admire, and things they like to do. This information is analyzed and put into a profile. Many standardized interest inventories are available through commercial companies and are used by high school and college counselors.

A limitation with this technique is that interest and aptitude are often confused. It is possible for students to have a high degree of interest in an area yet lack the aptitude to be successful in it.

Example

Interest Inventory

I Like School When . . .

Name_____ Date_____

I like school when . . .	yes	?	no
1. My best friend is in the class	☐	☐	☐
2. I can sit wherever I want to	☐	☐	☐
3. It's raining	☐	☐	☐
4. The sun is shining	☐	☐	☐
5. The teacher smiles at me	☐	☐	☐
6. I score high on a test	☐	☐	☐
7. We play games in class	☐	☐	☐
8. I meet new people	☐	☐	☐
9. I have time to sit and think	☐	☐	☐
10. I can keep busy	☐	☐	☐

Projective Technique

A projective technique allows students nearly complete freedom of response. Generally, these techniques require special training to administer and interpret. The ordinary classroom teacher may wish to use only the simple forms such as the open ended statements or unfinished sentences. Students may be asked to complete statements about themselves, school, family, friends, etc., such as:

When I baby sit . . .
My family is . . .
A true friend is . . .
This course has been . . .
In this course I expect to learn . . .
As a teacher you are . . .
On the job I . . .

Attitude Scale

If an objective in real life situations is to change student attitudes, any number of techniques, especially observations, can be used by teachers to evaluate change. If instead, we are concerned that students be aware of their own attitudes and the effect attitudes can have in their relationships with others, then self-report techniques can be employed. Attitude scales can

be designed to measure students' feelings toward some person, group, school, occupation or a specific course. This technique is generally most successful when students have least to gain by distorting the results.

Responses on attitude scales can be used to evaluate curricula, courses, teaching techniques, etc. The following statements are illustrative of those used on an attitude scale. The response pattern used below is one of several currently being used by commercial companies.

Directions: Circle the symbol that best represents your belief or feeling about the statements listed below. The responses are—strongly agree (SA), agree (A), undecided (U), disagree (D), and strongly disagree (SD).

SA	A	U	D	SD	1. This class is exciting.
SA	A	U	D	SD	2. I look forward to this class.
SA	A	U	D	SD	3. The teacher is willing to help me.
SA	A	U	D	SD	4. The teacher is fair.
SA	A	U	D	SD	5. What we are learning is important to me now.
SA	A	U	D	SD	6. What we are learning can help me in the future.

Information about the way one functions, feels and perceives oneself in real life situations can be helpful in making judgments about one's personal, social and emotional growth. These evaluative judgments are made with the intent of providing feedback to students and planning further experiences to enable them to reach a desired level of adjustment not for grading purposes.

Chapter 3

Simulation Techniques in Evaluation

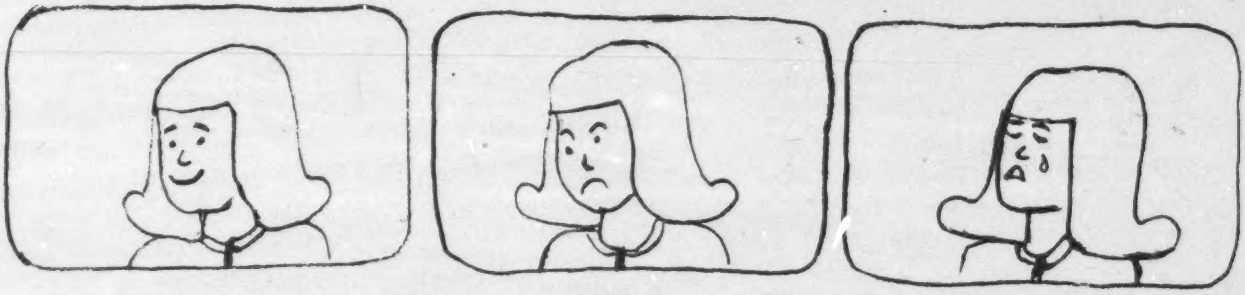
AS WE CHOOSE EVALUATION TECHNIQUES, we are trying to give each student an opportunity to show growth. Since students are all different and some perform well in one way, some in another, variety is needed to allow them all adequate opportunity to show what they have learned, whether they have changed, and how they have grown. If we have no opportunity to observe students in real life situations for evidence of learning, can we create situations to secure evidence, i.e., simulate reality, or use abstractions from reality? In this chapter we shall describe briefly, with examples, some possibilities for simulations. For the most part, we shall use the simulation techniques included in the earlier Home Economics Education Association volume, ***Choosing Techniques for Teaching and Learning***, Chapter 3.

As a working definition of an educational simulation, let us say that it is a pretense, a frankly faked situation contrived to represent some part of real life for educational purposes. Used for evaluation, this type of technique may be especially useful to the student who does not perform well on such abstract techniques as tests. (Unless we are evaluating a student's ability to take tests, there is little reason to use tests to evaluate a "poor test-taker's" learning in a particular area.) Some simulations have the advantage of not requiring reading and writing, and hence may be appropriate for those whose skills in these areas are minimal.

Simulations generally take more time than paper-pencil tests, and they cannot be machine scored. They would be used in combination with other techniques, of course, and they could secure certain kinds of evidence unobtainable in any other way. They can provide needed variety.

Role Playing

Role playing is one of those simulation techniques which need not require reading and writing. It provides the student opportunity to show behaviors and to incorporate principles and information learned, and it can be used as an evaluation technique in many situations.



Role playing may be done by individuals or by groups although the latter is more common. If a group is responsible, it may take very careful planning to evaluate a single individual. Sometimes it may be more effective to have a student direct the behavior of others involved or to role play alone if individual evaluations are desired. In such cases, however, a shy student may be at a particular disadvantage. S/he might be helped by having the teacher be a part of the role play. This could also give the teacher opportunity to create situations requiring responses that s/he wishes to evaluate in the student.

Role playing is a time-consuming technique even if, as is usually the case, each role play takes only three to five minutes. It will be only one of many techniques utilized and its use will be limited to evaluations that cannot be made as satisfactorily by other means. The teacher may select specific role playing situations for particular students or let each student "draw one from a hat" in a chance fashion. To get a double benefit from the time used, students observing the role play could evaluate, too. For example, in the example described below, the "audience" could be asked to "write a paragraph telling whether you think this consumer will be able to make a satisfying decision if s/he behaves as shown in the role play." The paragraph could be part of each student's total evaluation.

An Example:

Teaching objectives: (1) Students will know and use sources of consumer product information before making decisions involving money expenditure; and (2) students will shop around and compare prices and qualities before choosing consumer products and services. (These objectives involve the three highest cognitive levels: analysis, synthesis, and evaluation.)

Teacher asks students to "role play in three to five minutes a situation showing what you would do if you were setting up your own apartment, to live alone, and you needed to buy an iron." (The phrase, "to live alone," is included to eliminate the need for involvement of others in the decision.)

The teacher would need a scoring system and some kind of form on which to record observations made during the role play. The form could allow coding and thus minimize the time required to make the written record. The sample below is illustrative.

Class_____ Date_____

Student	Printed resources used	Other resources used	Market sources shopped	Notes
Sue				
John				
Dave				

Such a record would allow interpretation into a score if needed. It could allow a given number of points if the student showed use of such printed resources as Consumer Reports, or reliable human resources, or labels and guarantees. It might allow additional points if the student compared irons at different market sources. The number of points would depend upon the circumstances. The teacher would decide on these matters after thinking of such questions as: How does this fit in with additional evaluations? Should it be sophisticated enough to score differentially according to the number and dependability of sources consulted, the number of markets shopped, and the quality of the questions raised as sources and clerks were encountered? The teacher would also decide whether to allow preparation time, and if so how much, or to require immediate extemporaneous response in the role play.

Each role play assignment could be different so that one student does not simply repeat what another one has already done. Or, perhaps, only one or a few students would be evaluated via role play while the rest are using other means.

Role play and some of the other techniques that follow can be used in evaluating for affective objectives, too. Teachers nearly always have affective objectives along with the cognitive ones, regardless of the content being taught. They are trying to develop positive attitudes toward learning and toward the subject matter, to help students to greater self-esteem, and the like. The same role play may give evidence in both domains, but different observations would

be needed. In the affective area, the teacher would look for expressions of attitude, appreciation, opinion, and values. A different form would be needed for recording this type of information.

Skits and Pantomime

Skits and pantomime can be used in at least two ways in evaluation. A teacher-made or commercially available skit could be presented and the student asked to analyze it in some way, or the student could write his/her own dramatization to show understanding of learnings gained.

Example of Use of Teacher-made Skit in Evaluation:

Teaching objective: Student will recognize negative effect on a young child of the use of harsh punishment and negative commands. (This objective could involve cognitive levels 4, 5, and 6, i.e., analysis, syntheses, and evaluation.)

Teacher presents brief skit showing parent shouting "don'ts" to a two-year-old and then spanking hard.

Assignment for student: Tell whether this parent's behavior in this situation is similar to what you would do and, if so, why. If not, tell why not and describe the behavior you would substitute.

Possible scoring procedure: Points could be allowed for recognition of shouting, negative commands, and beating, and for explanation or description of the probable negative effect on the child and of the substitute behavior, if positive.

Example of Use of Student-made Skit in Evaluation:

Teaching objective: Student will be able to verbalize an alternative to fighting when employees have disputes. (Cognitive levels 4, 5, and 6)

Assignment for student: Write and enact a skit of less than five minutes to show understanding of "a peaceful way to settle a dispute between two employees on the job."

Possible scoring procedure: A five-point scale could be set up with descriptions for each level and student given five points "if skit shows settlement of dispute with no ill feelings and it looks like it will last" and one point "if situation in skit is little better than a fight." Descriptions for two, three, and four point would gradate from those given for one and five.

If each student skit is analyzed by all other students, then the benefits of both uses of skits in evaluation can be realized simultaneously.

Caution: To promote fairness, the points allotted on any evaluation device, including tests, should be in proportion to importance as related to objectives and considered along with all the others used in that grading period if they are to be summed for a total score. That is, there will be a reasonable relation between the number of points allowed for a given technique, or portion thereof, and the importance of what it is evaluating.

Educational Games

As a simple working definition of an educational game, we might say that it is a contest played by set rules with educational objectives. Some games are simulations (that is, what the players do in the game bears some relation to something they do in everyday life), and some are abstractions. And, alas, as Byrne has pointed out in *Games People Play*, some are real life situations! But here we shall be considering primarily the simulations.

Games can be used for evaluation as well as for teaching. If teams, rather than individuals, compete, then obtaining a score for each individual may be impossible unless team members take turns as individuals during play. For example, a teacher-made game similar to the TV jeopardy game would be devised using any desired subject matter.[1] Individual members of a team, in turn, could make the decision of which category of questions to try and what level of difficulty, then provide the answer, and score, if correct, for the team. Individuals could have the score on the answers they provided, and the team score could be the total of all its members.

Some games available from educational publishers can be played "solitaire" by one student in a manner that evaluates past learning.[2] And some, designed for teaching rather than evaluation, could be adapted for evaluation purposes. For example, a game that permits players to use reference material to look up answers while playing and learning could be played without the references to see what the student remembers. Scoring systems also need adaptation in such cases.

Evaluating by Inquiry

"Inquiry sessions"[3] can be conducted with individuals or with groups. They can be a sort of game in which students are shown a film, demonstration, or dramatization and inquirers ask their questions according to set rules. For example, a player could ask only questions that are answerable by Yes or No, and s/he could continue to question as long as s/he gets a Yes answer. The "winner" could be the first to explain the phenomenon observed and could obtain an agreed-upon score. If there is only one inquirer (player), the score could be based upon such

[1]For an example using insurance, see *Illinois Teacher*, Vol. XVIII, No. 2.
[2]An example is "The Calorie Game" from *Games That Teach*, 1107 West University Avenue, Urbana, Illinois 61801.
[3]See *Choosing Techniques for Teaching and Learning*, p. 20.

results as how many Yes answers s/he received, how many questions s/he asked before being able to explain the phenomenon, or the time required to explain it. The former would direct attention to the quality of individual questions, and the latter would promote speed with less reflection between questions.

Example of Evaluating with Inquiry Technique:

Teaching objective: Student will be able to explain three ways to remove stains from fabric. (Cognitive, level 4)

Teacher gives silent demonstration removing (1) oil stain from carpet by absorption with corn meal or starch and vacuum cleaner or whisk broom, (2) fresh blood stain from garment by dissolving in cold water, and (3) ink stain from table cover by bleaching with sodium hypochlorite solution. Then s/he asks student to name three general methods by which stains are removed.

Student inquires, with questions that can be answered by Yes or No, and then explains, if s/he can, the three methods of stain removal demonstrated: absorbing, dissolving, and bleaching.

Scoring could be decided in advance by teacher or student, or both together, according to the number of questions required to explain each or all three methods, or the amount of time required to solve.

Caution: In order for the above to be effective as evaluation in a given situation, it could not have been used previously for teaching. Simple memory and regurgitation of facts would suffice to solve, if the student had had the same experience earlier.

This is true for many, if not most, other evaluation techniques as well. If more than simple recall is being evaluated, the teaching procedures and the evaluation procedures must be different.

Demonstrations

The demonstration technique can be used in at least two ways in evaluation, in one case being performed by the teacher and in the other by the student. Given by the teacher, a demonstration, either silent or with verbal accompaniment, can be a prelude to an inquiry session, as noted above, or as a basis for students' written or oral analysis of what is seen. Given by the student, a demonstration can show understanding of principles learned or skills acquired.

Example of Evaluation by Student Analysis of Teacher Demonstration:

Teaching objective: Student will recognize that baking powder causes biscuits to rise.

Teacher gives silent demonstration of biscuit-making with deliberate omission of baking powder. As the biscuits go into the oven, teacher asks student to describe what s/he expects biscuits to look like as they come out of the oven. If there is only one student involved, description can be oral; if a group, it would have to be written in order for all to participate equally.

Possible scoring procedure: A given number of points (in relation to points allowed in other

accompanying evaluations) could be allowed those who answer correctly that biscuits will come out flat.

If students do not answer correctly, an inquiry session could be conducted when the biscuits come out of the oven, with the question being "Why didn't the biscuits rise?"

Example of Evaluation by Student Demonstration:

Teaching objective: Student will recognize safety hazards for young children in the home. (Cognitive, levels 4, 5, 6)

Assignment for student: Prepare a given space for young children to play and explain, orally or in writing, why it is safe. (Teacher may deliberately "plant" some hazards which student must remove.)

Possible scoring procedure: A given number of points could be allowed for removal of hazards and additional points for safeguards provided and for adequate explanations of why changes were made and why the space is now safe. Or, a scale with descriptions of several different levels of achievement could be prepared and the student rated at the level which best describes his/her achievement.

Other Examples:

To show understanding of principles of management in relation to home storage (cognitive, application), student could demonstrate rearranging some portion of the home, school, or office to increase efficiency and to explain why the changes were made. (Teacher could plan for inefficiencies that needed to be corrected if desired.)

To show skill in putting in a zipper (psychomotor, level 4-mechanism) student could demonstrate to less advanced student or an adult class or a visitor to the school.

Laboratory and Experimentation

We have often used laboratories in skill subjects to allow students to watch a teacher demonstration and then to repeat it at their own "lab stations" and thus to show whether they have acquired certain skills. In such cases, the teaching objective is usually psychomotor, guided response level.

Laboratories can also be used to give students opportunity to experiment, to do more than repeat what they have seen, step by step. They can be presented with a problem or an unknown situation and given the task of solving it by employing some principle they have learned earlier in a different situation. In this case we could be evaluating whether the student had learned a procedure for attacking a problem or whether s/he could apply a principle that was supposed to have been learned.

Beginning cooks could be given something to prepare without any directions or recipe. They could experiment with different ways to cook the food until something satisfactory is achieved and then explain why the final procedure was more satisfactory than the others. For example, they might be given a cereal like cream of wheat which would probably lump at

first or an egg which might be tough if cooked in certain ways. Experimenting with different procedures and applying principles learned earlier, such as, "high temperatures toughen protein" or "fine particles of starch lump in hot water unless separated in some way," would enable the student to create his/her own recipe or cooking directions and show that the principles were understood. In such a situation the teaching objectives would probably be cognitive 4, 5, and 6 and psychomotor 5 (complex overt response). There might also be affective objectives up to level 3 (valuing) if the teacher is trying to help the students to enjoy learning, to appreciate the procedures scientists use in their creative work, to commit themselves to investigation rather than blind acceptance of instructions, etc.

To avoid excess confusion and possibility of chaos, this type of experimentation might be used first with one or a few students to see if they and the teacher feel comfortable in the situation.

Possible scoring procedures could include evaluation of the product, the process, and the student's attitude. For the former, with a given amount of time in which to experiment, students could be given points or a rank on a five-point scale according to whether s/he produced an acceptable food, in terms of agreed-upon criteria. Or, with unlimited time in which to experiment, student could be scored on the basis of how long it took to prepare an acceptable product. In judging the process, the evaluator might look at the condition of the work area, the waste involved, the order in which procedures were tried and the explanations given, the evidence that student learned from each attempt and did not repeat mistakes, etc. In the affective area, the student's mood, enthusiasm, approach to the problem, attitude toward failures and successes, and the like, could be observed.

Caution: Laboratory work can be highly creative or very routine, depending upon whether student is simply following directions, recipes, and patterns or being given opportunity to try out own ideas.

If objective is to help student understand the scientific method, the way scientists operate in unearthing new knowledge, the routine, direction-following procedures will not suffice, even though they may sometimes be useful in reaching other objectives, e.g., to help student discover a principle.

Contrived Incident

A contrived incident[4] is not "frankly faked" at first. It pretends to be real for a time. As used by Raths et al.[5] it stressed the creation of a situation in which feelings led to new understanding.

As an evaluation technique, the contrived incident may place less emphasis on feelings, when used in the cognitive domain, and more on creating the need to show understanding previously gained. In fabricating the situation, the teacher may need collaborators. With this difference in emphasis, it may not be necessary to reveal the contrived nature of the incident and to discuss it immediately, as is so important when used in teaching via feelings. If, however, there is any chance for the student to discover the contrivance later, it would be advisable for the teacher to explain frankly and thus avoid a possible breakdown in trust.

If the contrived incident is used to evaluate in the affective domain, feelings may still predominate. The situation may be created for the express purpose of providing the opportunity, even the necessity, for the student to express a value.

Some examples may help to clarify.

---

[4]See *Choosing Techniques for Teaching and Learning*, p. 24.

[5]Louis E. Raths, Merrill Harmin, and Sidney B. Simon, *Values and Teaching* (Columbus, Ohio: Chas. E. Merrill Books, Inc., 1966), p. 123ff.

Example of Contrived Incident Used to Evaluate Extent of Knowledge Gained:

Teaching objective: Student will understand the adequacy, in terms of individual nutrients, of a vegetarian diet. (Cognitive 4, 6)

Teacher arranges for someone to interrupt class, either by telephone or in person, to ask "whether this vegetarian diet my teenager has gone on is going to ruin his health." Teacher asks a student to take the call or talk with the visitor in a quiet corner and offers to be a resource if needed. Teacher listens to conversation, or tape records it (with permission), or gets a report from the visitor later. Care would be needed in the latter case to avoid evaluating the student on the basis of possible error in the second-hand report. The visitor's understanding of what was said, however, is of crucial importance in the evaluation even if it is not the same as what the student said in trying to answer the question.

An alternative use of such an incident which might seem preferable to some teachers would be to present the problem to the whole class, let each student write some reactions to the question and then discuss it as a total group.

Possible scoring procedure: A ranking on a scale with descriptions for different levels could be assigned, or points could be allowed for the inclusion of such items as the difference between a strictly vegetarian diet and a lacto-ovo-vegetarian one, the complementarity of proteins, the absence of vitamin B_{12} in plant foods, the fiber value of many plant foods, and the like. With such a list and the opportunity to add to it, the student could probably score him/herself. To avoid dependence upon memory of what was said, the student could analyze a tape recording of the conversation.

Example of Contrived Incident to Evaluate the Development of a Skill:

Teaching objective: Student will show attainment of skill in the use of a washing machine and recognition of when its use is appropriate. (Cognitive, application, and psychomotor, unguided response or higher level depending upon the circumstances.)

Teacher brings to class a dress belonging to school secretary and says, "Miss X asked me whether this dress could be washed by machine. What do you think? She's coming in during her lunch hour. Who would like to answer her question and, if it is washable, show her how to do it in our washing machine?" (The chosen dress is washable but not obviously so.)

Possible scoring procedure: One could use a five-level scale with descriptions of each level, or allow points for such behaviors as examination of the label, observation for special non-washable features, justification for recommendation to wash, attention to water temperature, detergent used, length and gentleness of wash cycle, whether to remove before the spin cycle, how to dry, whether to iron, etc. Extra points could be allowed if dress owner understood explanations readily and/or for the volunteering to handle the problem.

Example of Contrived Incident Used to Determine Clarity of Values:

Teaching objective: Student will be able to verbalize own values in regard to use of time.

Teacher arranges for "an interviewer from one of the public opinion polls" to request an interview with a student and ask, "Will you rank the following in order of decreasing preference to you in the use of your time, and then add five other activities and show where they would fall in the ranking?"

Dancing
Watching a sport
Active participation in a sport
Learning about nature
Driving or riding in a car
Playing a musical instrument or singing
Studying
Listening to music
Working at a job for pay
Volunteer work
Gardening
Talking with friends
Walking, hiking, or climbing
Cooking
Eating
Sewing
Going to movies or plays
Writing
Reading
Going to parties
Bird watching
Doing chores at home
Camping
Buying clothes

Possible scoring procedure: Since the time required to respond to this question is probably an indication of the clarity of values held, the scoring might be based on the time used. Should such a score be a part of the total score used to determine a "grade"? If we are not judging the particular values held, but only their clarity and strength, and if the stated objective above is an acceptable one for the student, why not?

In-Basket

The "in-basket" technique, which can also be used for teaching, is a "natural" for evaluation. It also makes possible a little evaluation at a time on a continuing basis. Although many variations are possible, it could work like this:

Each student has his/her own "in-basket" (which could be a tote tray, drawer of a cupboard, folder in a file cabinet, or a box on a shelf) and each day for a given period (or each Monday and Thursday or whatever) the teacher places there an item which requires action for the student to get it into the "out-basket," i.e., the teacher's box or drawer. These could be the same for all students on a given day, the same for all students but given on different days, or differentiated for each student according to ability, interests, etc. A specified amount of time in class could be allowed for the student to act on his/her item, or the assignment could be to get it into the out-basket within 24 hours, depending upon the circumstances. The items could require written responses or they could involve other kinds of activity, or a combination of both. An example follows. Comments about scoring and teacher responses will be included with each step.

Teaching objective: Student will demonstrate ability to make and justify decisions in regard to selection and maintenance of clothing. (Cognitive 4, 5, 6)

Teacher will place in student's in-basket the following hypothetical situations:

Day 1: You have just received a birthday gift of $15 which the donor says is to be spent for clothes. Make a decision about how to spend it and explain why you chose as you did. Use catalogs here in the department and attach pictures and descriptions of your choice(s).

Scoring: If the student makes a decision of any kind, part of the teaching objective has been reached and points can be allowed. The other part of the objective, i.e., the justification of the decision, can be judged by the teacher as relevant or not, sufficient or not, etc., and points awarded accordingly. The judgment will not be affected by whether the choice itself was the same as the teacher would have made.

Day 2: One of the teachers here is willing to pay you $5 to let down the hem of her crease resistant polyester skirt. She wants to be sure that the change does not show when finished. Decide what you will do and report.

Scoring: If the student says simply "I'll do it," the teacher might return the in-basket item with the question: How are you going to keep the old hem-line from showing? Points might be scored for such things as recognition that it may be impossible to keep the old hem-line from showing in this fabric, comment on the reasonableness of the size of payment offered, appraisal of own skill level in relation to the task, time required to perform the task, etc. Extra points are in order if student is ingenious enough to suggest a way to hide the old hem-line with an attractive trim of some kind.

Day 3: (Or the following could be for four different days) The drama class needs your help on costuming the current play. There is no pay. Here are ways you could help: (1) advise them about how to iron the costumes in stock, (2) mend a costume (wool skirt) that has

a cigarette burn, (3) select from their costume inventory the most suitable attire for a secretary, or (4) help them improve their storage facilities for costumes so they require less maintenance and last longer. Decide what to do and report.

Scoring: If student says, "I don't have time to help," the decision must be accepted and, if justified, some score may be allowed. If student chooses to help, then points may be awarded according to the correctness of the advice (e.g., safe temperatures for ironing various fabrics, storage methods that minimize need to press and maximize protection from moths, workable mending procedures).

Day 4: Your best shoes got soaking wet and your white shirt spattered with mud when you had to walk home in the rain today. What are you going to do?

Scoring: Student could say, of course, that s/he would "give both to the Salvation Army or throw them in the trash can." If so, what could the teacher do but accept it?! To avoid such a problem, the teacher could phrase the question differently, e.g., what can you do to keep both in wearable condition? Then points could be allowed for workable procedures for removing mud stains, attention to the material of which shoes and shirt were made as decisions were determined for care needed, recognition that the sooner the care was given the more likely would be its success, or even suggestions about how the problem could have been avoided, such as pre-treatment of shoes to make them waterproof.

Day 5: The pants you bought yesterday (and have not worn) have a rip in a leg seam. What action will you take? Will this action be the same as it would be if the damage were a tear in the seat? Why or why not?

Scoring: Student might say "return the pants to the store" or, since the problem is so simple and could be repaired at home in less time than it would take to go to the store, simply sit down at the sewing machine and replace the stitching. Either would be worth a score. Answering the second question the student should show understanding that a tear is different from a rip and not easily repaired; hence, a suggestion of returning the garment to the store for refund or replacement should be worth a higher score than most other possible responses.

Day 6: Your new shirt, which was labeled *sanforized*, shrank so much in the laundry that you cannot wear it even though you followed directions for washing it. What will you do?

Scoring: To score on this one, student should show understanding of the term *sanforized* and recognize that the store should "make it good."

Discussion

Teachers often use discussions as a teaching technique, and perhaps just as often some think they are using discussion when, in fact, it is something else, e.g., recitation.[6] Whereas, in recitation the teacher asks questions and the students supply answers on matters of fact, in discussion the students, and often the teacher too, offer expressions of opinion.

Discussion can also be used in evaluation. In some cases it could be tape-recorded (after students are accustomed to the machine and are not inhibited by it!) and analyzed later. It may be unrealistic to expect that a teacher will have time to listen to the discussion again on tape, but playing it back for the class for student analysis and evaluation may sometimes be appropriate.

---

[6]See "Discussion: An Oft-Abused Potential for Learning," by Connie Sasse in *Illinois Teacher*, Vol. XVI, No. 4.

Possible purposes for using discussion in evaluation may be suggested by the following questions:

Did all members of the group participate? With what frequency?

What was the quality of each contribution? Was it relevant? Did it show depth of thinking or feeling? Did it further or stifle the discussion? Did it create any hard feelings? Develop self-esteem?

Did any attitudes change during or following the discussion?

Did values get expressed? Was there tolerance of differing values?

Did the group arrive at consensus on action to be taken?

One of the questions to ask oneself in the use of discussion is whether the subject being considered is a "discussable" one. It is not appropriate, for example, to "discuss" the nutrients contained in milk since it is not a matter of opinion. Milk *does* contain certain nutrients. One *could* discuss whether society has a responsibility to see that all babies have enough milk for their optimum growth.

Another problem in the use of discussion in evaluation is finding an appropriate means of "recording the evidence." Still another is that of leading the discussion and recording the evidence at the same time. Sometimes a student or invited leader could provide help on one of these tasks.

One kind of form which might facilitate the recording of evaluations from discussion is shown below. Having a space for notes may enable the teacher to write enough to recall to mind later a contribution of a student which needs to be followed up with a conference or which could help explain an evaluation in another column. The categories shown are only some of the many possible ones.

Class_____ Date_____

Student	Frequency of Contribution	Quality of Contribution	Relation to Fact	Relation to Values	Notes
Sue	1	R	1	Exp	
John	~~HH~~	U	2	Con	
Dave	11	I	3	Int	
Karen	1	D	4	No	

Possible codes:

R—relevant
I—irrelevant
U—unusual, creative
D—destructive

1—used facts
2—based on fallacy
3—no facts involved
4—synthesized facts

Exp—expressed values
Con—showed conflicting values
Int—intolerant of values of others
No—no values expressed

The above is not meant to be a scoring device but only a vehicle for recording observations during a discussion in an abbreviated way. Hence, the code systems are not arranged consecutively from "good to poor" and are not intended to be hierarchical. The teacher can devise his/her own abbreviations to suit individual needs.

Noting frequency of contribution to discussion does not mean that high frequency means high grades. It does provide information which may help a teacher to recognize those who tend to dominate a discussion and those who are too shy to speak. For the latter there is a special need to provide other means of evaluation where shyness is not inhibiting.

Q-Sort

The research technique known as Q-sort can be adapted for evaluation purposes in both the cognitive and affective areas. In using it, a number of cards are prepared, each with a different word or phrase, and the student is asked to sort them into a specified number of piles, with a specified number in each pile, according to some given criterion. This activity may be more interesting to some students than taking a test even if the same material is covered. The card may contain pictures or samples rather than words. Some examples may clarify.

Example of Q-sort Used to Evaluate Knowledge Gained:

Teaching objective: Student will identify the level of vitamin A present in given fruits and vegetables. (This objective would be cognitive level 1 if the material has been taught and memorized but level 3—application—if general principles have been taught which can be applied in this situation.)

Student is given pictures of the foods below, or cards with names of the foods, and asked to sort them into three piles according to their level of vitamin A. Pile 1 should contain *13* foods which have little or no vitamin A, pile 2 should contain *8* foods which have a moderate amount of vitamin A, and pile 3 should contain *8* foods which have a great deal of vitamin A. Foods to include: apples, apricots, asparagus, dry beans, beets, cabbage, cantaloupe, carrots, cherries, red grapefruit, white grapefruit, grapes, green peas, oranges, peaches, pears, pineapple, plums, potatoes, sweet potatoes, prunes, raisins, spinach, dark squash, strawberries, tomatoes, turnips, turnip greens, watermelon. The number to be placed in each pile is determined by the levels of vitamin A in the foods. If more exactness is desired, pile 1 could be designated as foods having less than 8% RDA of vitamin A, pile 2 as foods having 12 to 30% RDA and pile 3 as foods having 45% RDA or more.

Scoring could be done by counting the foods in the correct piles and subtracting those in the wrong piles.

Example of Q-sort Used to Indicate Values:

Teaching objective: Student will be aware of own values in regard to clothing selection. (Affective 3 and 4)

Student is given cards with words indicating factors to consider in selecting clothing with instructions to sort into 3 equal piles labeled: Factors generally very important to me in making decisions about what clothes I buy, factors of some importance, and factors of little or no importance. Such factors as the following might be used: color, size, cost, washability, durability, fit, fashion, style, amount of care needed, cost of care needed, storage requirements, relation to clothes already owned.

Scoring: Since there are no right and wrong answers in this situation, the evaluation could only be in terms of whether the student was able to decide what factors were important to him/her and how long it took to make the decisions. The advantages and disadvantages of quick decisions could be discussed and implications understood.

Other Examples:[7]

Samples of fabric might be sorted into piles according to which could be laundered in the same washer loads.

Pictures of consumer products might be sorted into piles according to student preference of how to spend a given amount of money.

---

[7]An example of the use of the Q-sort in factors for selecting an occupation is given by Margaret Barkley in *Illinois Teacher* (March/April 1975), pp. 206-7.

Related Techniques:

If student objects in the clothing example above that "*all* these factors are very important to me," then they could be ranked instead of labeling *some as of little or no importance, etc.* Ranking can be applied in many situations in both cognitive and affective areas. If used in the affective domain, a cautionary note is in order. Trying to place values in a general hierarchy may be impossible and can lead to misunderstanding. For example, in using values to select housing one can determine which value predominates over another value only in relation to a given house. Trying to place a series of values in a rank order and then applying this to a housing decision disregards the degree to which each value is held and how individual houses might meet value needs in differing ways.

Another technique is "choosing from pairs" according to preference or according to fact, e.g., which of these two foods has more calcium?

Writing Letters and Filling Out Forms

Another simulation of a real life situation which may be useful in evaluation is filling out forms—income tax forms, job applications, questionnaires, credit applications.

Letter writing is a part of everyday life that can also be simulated. Evaluations can be obtained from "pretend" business letters, letters of inquiry, consumer complaint letters to companies or appropriate governmental agencies, or letters to family and friends in which certain content is specified, e.g., "write a letter to your pregnant cousin and advise her about how to take care of herself and her future baby." Another possibility is a reply to an Ann Landers-type letter which the teacher or other students may write to elicit knowledge and/or opinions in specified areas.

Audio-Visual Materials

Audio-visual materials may be useful in many simulations and sometimes in other types of evaluations too. Cross[8] has given a discussion of several of the common ones, such as tape recordings, films and filmstrips, photographs, transparencies, graphs, pictures, and cartoons.

[8]Aleene Cross, *Home Economics Evaluation* (Columbus, Ohio: Chas. E. Merrill Publishing Co., 1973), ch. 13.

Chapter 4

Evaluating with Tests

EVALUATIVE DATA OBTAINED FROM TESTS can be used to (1) diagnose student learning disabilities, (2) make curriculum decisions and (3) make judgments about student academic growth and achievement.

Tests can provide some types of evaluative feedback that other techniques cannot provide. The decision of whether and when to use tests is determined by the kind of feedback that one wishes to obtain and the objectives the learner set out to accomplish. A comprehensive evaluation program requires the use of a variety of techniques to determine effectively the degree to which learning is being achieved.

The degree to which one can rely on the results obtained from written, oral, and performance tests, is in part determined by the quality of the tests.

The aim of this chapter is to provide some suggested guidelines for developing written tests that will yield results which possess content validity and reliability.

Written Tests

Written tests are one means of determining the degree to which certain objectives have been achieved but are totally inappropriate in some cases. They are more appropriate for cognitive objectives than they are for affective or psychomotor type objectives. Knowledge and intellectual abilities and skills are often evaluated on the basis of results from paper-and-pencil tests, although there are other satisfactory techniques for doing so.

Written tests can be classified as objective tests or essay tests. The term objective refers to the method of scoring the test items. *Objective tests* require the student to respond with a one or two-word answer, or to select the appropriate answers from among several alternatives and can be used to measure achievement at all cognitive levels. The scoring of objective tests can be quick, consistent and accurate because there is generally one correct answer to each test question.

Essay tests consist of questions or items for which the student is asked to prepare an extensive, organized written response of a paragraph or usually more. The accuracy and quality of the response is judged subjectively by the teacher. Essay test items can be written to measure achievement at all cognitive levels including synthesis and evaluation.

The appropriateness of each type of test depends upon the learning objectives being measured because they each provide different kinds of evidence of student achievement.

Constructing Written Tests

Tests constructed by classroom teachers are more commonly used than standardized tests. Teachers, therefore, need to plan tests to insure a high degree of content validity by identifying objectives that can be measured by a paper-and-pencil test and determining the relative importance of the objectives. The teacher can construct test items, either objective or essay type, consistent with the subject matter content and the behaviors identified in the objectives.

Table of Specifications

A table of specifications is a systematic procedure for constructing a test based on content studied and the behavioral outcomes stated in the objectives. The behavioral outcomes can be classified into many categories. Often the six major categories, or some portion thereof, of the taxonomy of educational objectives for the cognitive domain are used.

As an illustration, consider a segment of a course in child development that has the following learning objectives:

1. Increased ability to explain some aspects of the physical growth pattern from conception to early childhood. (Comprehension)
2. Increased understanding of nutritional needs from conception to early childhood. (Comprehension)
3. Increased ability to plan, prepare and serve a balanced meal for young children. (Application and Synthesis)
4. Increased ability to appraise the daily diet of young children. (Evaluation)

Using these objectives the teacher can identify the behavioral outcomes, and areas of course content, and indicate the relative importance of each as percentages on the table of specifications. The following illustrates one possible decision that could be reached about the segment of the child development course:

Table of Specifications

Behavioral Outcomes	*Areas of Content*
1. Knowledge	1. Physical Growth Pattern
2. Comprehension	2. Nutrition
3. Application	3. Meal Planning
4. Analysis	
5. Synthesis	
6. Evaluation	

Table of Specifications for a Segment of Course in Child Development

Behavioral Outcomes

Areas of Course Content	Knowledge	Comprehension	Application	Analysis & Synthesis	Evaluation	Total
Physical Growth Pattern	5%	5%	10%	10%	0	30%
Nutrition	5%	5%	10%	5%	25%	50%
Meal Planning	0	5%	0	10%	5%	20%
Total	10%	15%	20%	25%	30%	100%

These percentages can be used as approximations of the percent of test items that will be based on content areas and behavioral outcomes. The minimum number of items in the above test would be 20.

Sampling

Test items consist of samples of all the items that could be constructed because it would be impossible to ask every conceivable question on a test. Teachers assume that the samples chosen are representative enough to permit all students to demonstrate their individual levels of achievement accurately. Tests should include some simple (easy) items for motivation as well as some complex (difficult) items to provide a basis for identifying the higher levels of achievement. For motivational purposes it is better to put the easy questions first and gradually move into the more difficult items.

The content validity of teacher made achievement tests can be increased by (1) using enough test items to sample all the content learned in a given length of time, (2) including content based on all course objectives, (3) avoiding ambiguously stated test items, and (4) avoiding questions based on isolated and obscure facts. Often test items that are taken directly from a textbook become ambiguous when read out of context and involve pure recall.

In general, the longer the test the higher the reliability because scores are not likely to be distorted by students guessing correctly on all questions, and increased length tends to mean greater representation of content. Teachers need to be cognizant of the number of questions that students can answer in a given period of time.

Supply-Type Objective Items

1. Illustrations of Short Answer Items:

   Directions: In the blanks provided, write the word or words that correctly answer the question.

   What is the name of a nutrient that helps the eyes adjust to the dark? _____

   What is the hourly wage of a person who worked three hours and earned $4.50? _____

2. Illustrations of Completion Items:

   Directions: In the blanks provided, write the word or words that correctly complete the sentence.

   If a loan company charges a monthly interest rate of 1½%, then its annual interest rate is _____%.

   The primary colors are _____, _____ and _____.

Suggestions for Constructing Supply Type Objective Items

1. Test items in which the answer is definite and limited to a word, phrase, number or symbol tend to be less confusing than those with several blanks.

Example:

Poor: Foods that provide little or no _____ _, _____, or _____ are called _____ _____ foods.

Better: Empty calorie foods are those that contain little or none of the nutrients (*protein*), (*vitamins*), and (*minerals*).

[1]For an in-depth discussion of the writing of test items, see Robert L. Thorndike, ed., *Educational Measurement*, 2nd ed. (Washington, D.C.: American Council on Education, 1971), ch. 4 and 10.

2. Statements taken directly from a textbook are often confusing when read out of context and often several possible answers could be correct.

Example:

Poor: George Washington was _____ (answer could be any of these: a man, a soldier, an American, first President of the United States)

Better: The first president of the United States was _____. (*George Washington*)

3. Direct questions rather than incomplete statements generally provide greater structure and clarity.

Example:

Incomplete Statement: The difference in amount between three decigrams and one gram is _____ decigrams.

Direct Question: How many decigrams are equivalent to one gram? _____

Selection-Type Objective Items

Response Variations to True-False Test Items

1. Directions: If the statements below are true write "T" in the blank *after* each number, if false write "F" and change it to make it true. If you do not know, write "O."

   In the blank *before* each number, write a number to indicate your degree of certainty of your answer as follows:

   1. I'm not at all sure about this.
   2. I'm somewhat certain.
   3. I'm very certain about this.

   Your score will be determined by adding the numbers in the left blank for those that are correct and subtracting the numbers for those that are wrong.

   ___ 1. ___ Skim milk is approximately equal to buttermilk in nutritive value.
   ___ 2. ___ One form of anemia can be caused by an insufficient amount of iron in the diet.

2. Directions: If the statements below are true write "+" in the space provided, if false write "0." Circle the underlined word or words that make the item false.

   _____1. Tangerines, grapes, grapefruits and oranges are citrus fruits.

   _____2. Thiamin, riboflavin, iron and niacin are B vitamins.

3. Directions: If the statements below are true, write "T" in the space provided, if false, write "F," and if opinion write "O."

   _____1. Everyone needs to own a life insurance policy.
   _____2. Life insurance premiums vary in cost.
   _____3. The premium rates on group health insurance policies tend to be higher than on individual policies.

Suggestions for Constructing True-False Items

1. Some words (such as always, never, all, none, or only) tend to make items false, while others (such as often, usually, generally and sometimes) tend to make items true. Hence students may be given external clues to the correct answers.

Example:

1. All banks charge higher rates of interest than credit unions.
2. Finance companies generally charge higher rates of interest than banks.

2. Students' learning styles are generally affected by the way in which they expect to be evaluated. Therefore, the use of test items based on the memorization of trivia such as specific dates, names and places may encourage students to learn such to the exclusion of more important content.
3. When an equal number of true and false items are listed in alternate order or in some pattern, students may be able to guess the correct answers.
4. Opinions and prescriptive statements cannot be judged true or false. One can, however, ask students to identify the known opinions of prominent people, organizations, etc.

Example:

T or F

_____ Some opponents of the Equal Rights Amendment believe that it will force all women to enter the labor force.

_____ Ralph Nader believes that Congress should enact more stringent regulations to protect consumer rights.

Multiple Choice Test Items

A multiple choice item consists of a problem and suggested solutions. The problem is called the *stem* and the suggested solutions are called alternatives. The correct alternative or option is the answer and the other alternatives are called distracters or foils. Multiple choice test items can be designed to measure student achievement at all cognitive levels. The stem of a multiple choice item can be stated as a direct question or as an incomplete statement.

Example:

Teaching Objective: Students will be able to identify major food sources of nutrients.

Directions: For each of the following questions choose the correct answer from among the four alternatives listed. Write the letter opposite your choice in the blank provided.

_____ Which of the following fruits provides the greatest amount of vitamin A?

A. one fresh yellow peach
B. half cup of fresh strawberries
C. one fresh sliced banana
D. half cup of raisins

Note: A variation of the multiple choice test is to ask the student to identify the best answer rather than the correct answer among the alternatives and justify the choice.

Suggestions for Constructing Multiple Choice Tests

1. When a concise and meaningful problem is stated in the stem of the item, greater clarity is provided than when the stem is very brief.

Example:

Poor: Chicago is
A. the largest city in Illinois
B. the largest in the southern tip of Illinois
C. surrounded by water
D. governed by a city manager form of government.

Better: What is the largest city in Illinois?
A. Peoria
B. Chicago
C. Rockford
D. Joliet

2. When all of the alternatives are not grammatically consistent with the stem, it may be easy for students to guess the correct answer.
3. When alternatives are plausible and homogeneous with regard to the learning being measured, uninformed students are not provided clues to the correct answer.

Example:

Which of the following measurements is equivalent to one meter?

Poor Alternatives:	Better Alternatives:
A. 2 yards	A. 100 centimeters
B. 30 inches	B. 150 centimeters
C. 150 millimeters	C. 10 centimeters
D. 100 centimeters	D. 1000 centimeters

4. When the alternatives are randomly ordered, the correct answer is not likely to appear in the same position for each item.

Matching Test Items

Matching test items consist of two lists or columns of items with instructions for matching each of the two lists. The items in the list for which a match is sought are called *premises*, and the items in the list from which the selection is made are called *responses*. Matching items can test students' ability to identify relationships or to make associations to a limited extent.

Example:

Teaching Objectives: Students will be able to comprehend metric terminology.

Directions: Match the metric measurement term with the type of measure for which it is used. Write the letter in the blank.

Metric Measurements	*Common Types of Measurements*	
_____ 1. gram	A. length	D. volume
_____ 2. metre	B. mass	E. temperature
_____ 3. liter	C. strength	

Note: Variations in matching test items can take several forms. (1) Imperfect matching test items have a shorter list of responses than premises. Sometimes a response may be used more than once or not at all. (2) Compound matching items consist of one list of responses and two lists of premises.

Suggestions for Constructing Matching Tests

1. The use of homogeneous material in each matching item reduces the probability that students will correctly guess the answer.
2. Student confusion is less likely to occur when the basis for matching the responses and the premises is clearly stated.
3. Students are less likely to overlook parts of a matching item when they are all on the same page.
4. The use of an unequal number of responses and premises, and responses that can be used once, more than once, or not at all, reduces the probability of correctly guessing the answer.

Preparing and Scoring Essay Test Items

Measurement specialists classify essay items as *extended response* or *restricted response* depending on the amount of freedom the student is permitted in his/her response.

The type of achievement that one wishes to measure will determine whether essay test questions should be used instead of objective items. Essay tests can provide data about a student's ability to organize and present information in a logical manner. Essay questions enable teachers to collect evidence about achievement as well as the student's ability to communicate in written form. The scoring of essay tests can be highly subjective; therefore, teachers may wish to guard against this potential weakness.

Extended Response Item

The extended response item gives the student a great deal of freedom and flexibility to organize his response to include all information that he deems pertinent and relevant. It is possible that an extended response type item can be so constructed that a student could demonstrate the achievement of knowledge, comprehension and the ability to apply, analyze, synthesize and evaluate facts or ideas. Extended responses can also reflect student attitudes, interests and creative abilities. An example of an extended response type item is shown below:

> Mary wants to make a dress. The fabric she purchased is 100% cotton, and the label on the fabric bolt did not indicate that it was preshrunk. Her thread is polyester, the fabric on the zipper is preshrunk. Should Mary make the dress slightly larger than her size to allow for shrinkage or shrink the fabric before making the dress? Why? What would be the likely consequences of each procedure?

Restricted Response Item

The restricted response type item defines the way the student is to respond to a specific problem. It requires the student to recall information, organize it in a logical manner, arrive at a conclusion, and state it in a limited number of sentences or paragraphs. These restrictions simplify the scoring and improve its reliability. As an example, see the restricted response item below:

> What is the meaning of the statement, "You are what you eat"?

Suggestions for Constructing Essay Test Items

1. Confusion is less likely to occur when the questions are stated concisely and the students' task is specified clearly.
2. If students have the choice of selecting a certain number of essay items out of the total, they may be taking different tests and the achievement being evaluated may not be comparable.
3. A teacher may wish to determine the approximate amount of time needed to answer each question and suggest a time limit for each question. Students often are penalized for working too slowly.

Scoring Essay Test Items

The scoring of essay test items can be made less subjective and more reliable when:

1. a scoring key for evaluating students' answers indicates the major points that should be included in the answer, the expected sequence or organization, and determines the amount of credit or range of credit that will be allotted each item. If papers are read before key is made, student answers can be considered in formulating the key.
2. one is aware that factors such as handwriting, spelling, punctuation, neatness, etc., may influence scoring. The degree to which these factors should be reflected in the score depends on the achievement being measured, and students need to be informed of the criteria for evaluating responses prior to administering the test.
3. all of the answers to one question are read and scored before going on to the next question. This lessens the possibility of the teachers' standards or expectations changing as a result of interruption of thought pattern.

4. the identity of the students is unknown until all of the papers are scored. Teachers are less likely to be influenced by past behavior.

Performance Tests

Performance tests require the student to manipulate objects or perform tasks for someone to evaluate the degree to which a skill has been achieved. The achievement of psychomotor objectives can be measured with performance tests. Poor readers, writers and students with language problems may be able to demonstrate a higher level of achievement on performance tests than on written tests.

Standards for evaluating student performance need to be identified prior to the testing situation. The teacher can determine the procedure to be followed to complete the task, describe what the product of each completed step will be, and determine the amount of credit to be given for each step and the total possible score to be earned for successfully completing the entire task. There may be a range of credit depending on the quality of the performance. Some steps may be more important to the success of the final product than others and should, therefore, be worthy of more credit when scoring.

When performance tests are used, the *procedure* used and the final *product* can both be evaluated.

Oral Tests

Oral testing is used far more often in higher education than secondary education. It may, however, be feasible for a teacher to make the option of oral testing available to students who for some reason cannot take a written test.

A major advantage of oral achievement testing is that when a student commits an error, the teacher can probe to reveal the causes of the error. Oral testing can be very time consuming and unless a tape recorder is used, there is no record of the student's answers. Oral testing can create a higher level of student anxiety than written testing for some students. Teachers, therefore, need to reduce the student's anxiety level by using an encouraging tone of voice, briefly talking with the student to put him at ease prior to asking test questions, and meeting in a quiet and pleasant environment. It is also important to permit students enough time to compose their answers, and to leave a question and return to it later in the testing period as is done during written tests.

Other Written Exercises

The use of some types of puzzles for evaluation purposes is unfortunately becoming quite common. Student reaction is often mixed. Some students enjoy the challenge of putting clues together and solving puzzles and others don't. Unlike taking written tests which most students generally associate with school, solving puzzles is an exercise often viewed negatively.

Crossword puzzles and word scrambles are more commonly used than jigsaw type puzzles. They are generally used to evaluate the achievement of objectives at the lowest cognitive level. The ability to recall definitions, abbreviations, names, places, etc., is the behavior most often measured by crossword puzzles. Puzzles that require students to find certain words among scrambled alphabets have little value as an evaluation technique.

Crossword puzzles and word scrambles may be more helpful as a technique for teaching than for evaluation because some students may enjoy the challenge of completing the puzzle as they learn to define words. However, if one is trying to determine whether students can define terms, other forms of written tests can be more direct, understandable, and less time consuming.

Chapter 5

"Putting It All Together"

PUTTING IT ALL TOGETHER is the hard part of evaluation. Is it necessary to put it all together? When, in life outside of school, are evaluations "put together"? Nowhere except in school does anyone try to abbreviate all evaluations into one letter, A, B, or C, but there are real life situations in which a series of evaluations are looked at together to determine whether one will receive a promotion, a pay raise, an "efficiency rating," or an award of some kind. In some of these cases it is a dichotomy—you get the promotion or the award, or you don't—and in others a continuum—you get a little raise or a larger one, a moderate rating or a higher one.

Hence, cumulating many evaluations into one, difficult though it is, has some value. It also has some disvalues, as has been pointed out by such educators as Simon, Johnson, and Melby.

Kirschenbaum, Napier, and Simon[1] suggest that grading creates anxiety, interferes with learning because the grade rather than the learning becomes the goal, encourages conformity because fear of low grades prevents students from stating creative and non-conforming ideas, and increases competition.

Johnson's[2] research points out the disadvantages of competition when he says that "cooperative learning structures, as compared with traditional competitive learning structures, will increase student achievement in problem solving tasks, increase positive interactions among students, increase positive attitudes toward learning and school, increase values of egalitarianism and appreciation for differences among individuals, and decrease anxiety about learning."

Melby[3] uses strong words to speak about grading. "Our marking system," he says, "is no longer relevant to the needs and educational programs of our society. It says nothing meaningful about a pupil. . . .It leads us to measure the outcomes of our educational programs in terms of what people know, when we ought to be measuring them in terms of what people are and are in process of becoming. . . .the marking system is not only irrelevant and mischievous, it is destructive. It destroys the self-concepts of millions of children every year." Melby believes that even those who receive high grades are injured "because they are induced to seek the wrong goals." He continues, "When you go to the doctor for a history and physical check-up, the doctor writes constantly as you talk describing your condition. He accumulates for you a medical history. He does not give you an A or a B (in health). He does not use meaningless terms which blot out your individuality."

Sometimes the grading system creates an adversary relationship between teacher and students. The external examination system, though it has little else in its favor in our judgment, eliminates this problem as teacher prepares student to take the exam.

Regardless of how one stands on the issue of putting it all together in one grade, the fact remains that at this point in time many thousands of teachers are in positions of no choice. They have to "give grades." While they are working to change the system, if they believe it needs to be changed, they long for help on how to be fair in assigning grades.

---

[1]Howard Kirschenbaum, Rodney Napier, and Sidney Simon, *Wad-ja-get?* (New York: Hart Publishing Co., Inc., 1971), pp. 121, 129.

[2]David W. Johnson, "Affective Outcomes," in *Evaluating Educational Performance*, ed. by Herbert J. Walberg (Berkeley, Calif.: McCutchan Publishing Corp., 1974), pp. 104-5.

[3]Ernest O. Melby, "The Deprived Child: His Gift to Education," in *The Community School and Its Administration*, Vol. IV, No. 12 (August 1966).

We have no recipes or patterns that guarantee success. We can only offer some ideas to think about and some alternatives to try. Looking back at some of the principles of evaluation on page 4, we can find some food for thought. Our thinking and experience have led us to believe that grading tends to be more just, fair, and equitable

- when many kinds of techniques are used to gather the evidence
- when written records are kept
- when evidence is gathered over a considerable time span
- when more than one person is involved in making the judgments
- when many kinds of behaviors are considered relevant
- when objectives are clear but not inflexible
- when human relationships are positive
- when there is an absence of prejudices regarding race, sex, socioeconomic level, etc.

One way to sum many evaluations into one is simply—though it is not simple at all!—to put a numerical score on each one and total the scores. If there are many scores, the error of reliability on one may be compensated by an error in the other direction on the next, as occurs in random sampling. One of the big tasks in using this procedure is to avoid having more points on any single item in the evaluation than its importance to the total may warrant.

Another way is to use symbols (letters, pluses and minuses, numbers 1-5, or whatever) to place an evaluation on each item and then to average them all for a final grade. Again, there is danger that less important items may receive equal weight with more important ones.

Still another procedure, which overcomes some of the disadvantages of the two above, is to use a series of rating scales for different items and allow varying weights according to their importance. The items to be included and the weight to be assigned to each can be determined cooperatively by teacher and students. There could even be some variation between students, if desired, so that substitutions are possible on some items to allow for individual differences.

Determining Grades

Before one can confidently combine all evaluative data into a composite grade, several decisions must be reached.

1. What evaluative evidence will be reflected in the final grade?
2. Will grades depict the amount of growth or the final level of achievement?
3. What standards will each letter grade represent?
4. Considering the total objectives, what is the relative value or importance of all of the individual evaluative data for each student?

What Data to Include

Regarding the first question above, teachers should realize that it is not always necessary for every bit of evaluative data to be reflected in the final grade. For example, data about students' personal and social development can be very useful to a teacher; however, it is almost impossible to reduce such data to a letter grade. Such data will be far more useful and meaningful when communicated in descriptive terms to students and parents for the purpose of assessing present development in order to plan for future growth.

Some data may have been collected to determine the effectiveness of certain teaching techniques or to determine whether curriculum revisions were necessary, in which case, the evaluation instrument may have been designed for these purposes rather than to measure achievement.

What evidence should be reflected in the grade? All evidence that is directly related to the instructional objectives that one set out to accomplish should be reflected in the grade.

The form below is illustrative of this third procedure. It can be used in all domains—cognitive, affective, and psychomotor—separately or together.

Directions for using the form:

Fill in at left as many items as you wish to consider for the total evaluation. Write descriptive terms under the numbers 1 to 5 for each item. Rate the items on each line by circling the number with the description nearest to level reached. Indicate importance of each item by adding a number in the column after the X. Multiply the circled number by this weight and place in right hand column. Total right hand column and translate into grade.

The items in the left hand column can be techniques and devices used to secure evidence (e.g., role play, demonstration, exam, discussion, etc.) or abilities to be evaluated (e.g., manipulative skill, verbal fluency) or types of content (e.g., nutrition, consumer competence, management principles).

EVALUATION SUMMARY

Name_____ Date_____

Course_____

	1 2 3 4 5			
1.	_____	X	_____ Weight	= _____
2.	_____	X	_____ Weight	= _____
3.	_____	X	_____ Weight	= _____
4.	_____	X	_____ Weight	= _____
5.	_____	X	_____ Weight	= _____
6.	_____	X	_____ Weight	= _____
7.	_____	X	_____ Weight	= _____
8.	_____	X	_____ Weight	= _____
9.	_____	X	_____ Weight	= _____
10.	_____	X	_____ Weight	= _____

Etc. Total_____

Recommended final grade_____

Grades as Evidence of Growth or Achievement

Will grades depict the amount of growth or the final level of achievement, or both? Ahmann and Glock[4] state that these are two different concepts and make the following distinction. Growth is personal academic gain. The indication of growth is based on pre-instruction and post-instruction measurements. Growth also takes into consideration the individual, his ability, prior experience, present conditions, etc. On the other hand, achievement means the student's present academic status which is generally judged against norms or teacher standards.

Differences between the two concepts can be described by use of a diagram. (Data are hypothetical but could have been collected by such instruments as Spitze's Nutrition Knowledge Test for Consumers.[5])

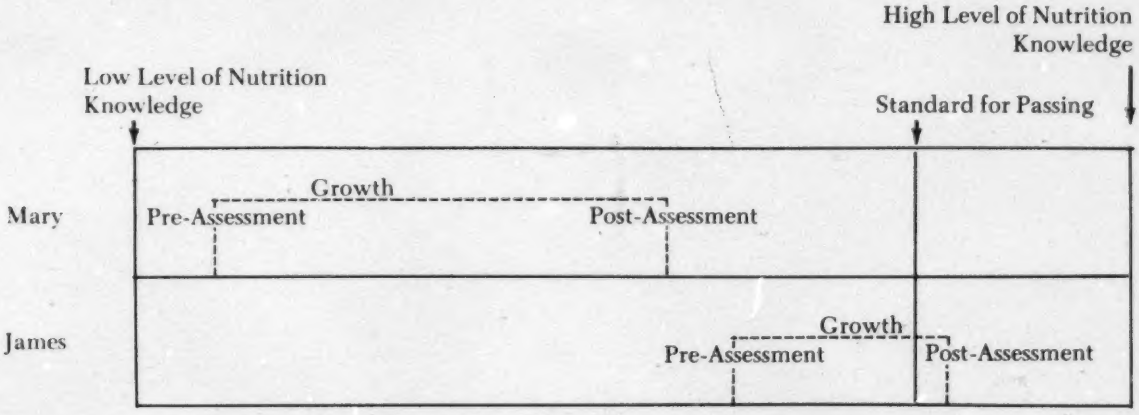


Diagram of Growth and Achievement in Nutrition Knowledge
(Pre- and Post-Assessment)

It is possible for a student to show evidence of growth in a course based on pre- and post-assessments and still not achieve the standards set for passing a course. In the diagram above, Mary's overall gain was greater than James', but she did not achieve the standard set for passing the course.

Evidence of both growth and present status can be useful to a teacher in determining a final grade. However, Ahmann and Glock[6] contend that present status rather than growth should be the determining factor for several reasons. First, data on growth or gains tend to be unreliable because they include errors from both pre- and post-assessments. Second, if the pre-assessment instrument is too easy, and able students score extremely high, they will have little opportunity to improve on the final assessment. Third, it is harder to raise a high score even higher than to raise a low score.

Some students are very discouraged about school because they always receive low grades. It may be psychologically sound in some instances to permit such students to compete against themselves and base their grades on growth to avoid having them fail.

---

[4]J. Stanley Ahmann and Marvin D. Glock, *Evaluating Pupil Growth: Principles of Tests and Measurements*, 5th ed. (Boston: Allyn and Bacon, 1975), p. 371.

[5]Hazel Taylor Spitze, "Nutrition Knowledge Test for Consumers," Division of Home Economics Education, 351 College of Education, University of Illinois, Urbana-Champaign.

[6]*Op. cit.*, p. 502.

Standards for Letter Grades

Each teacher seems to have some notion of the standards of achievement which s/he can equate with a letter grade. The major problem is that too often the standards vary so greatly that a uniform interpretation of grades from several teachers is impossible.

It is the policy of some school districts to issue a single grade that reflects both student achievement and citizenship. It is possible under such circumstances for a student who has achieved well academically but has a poor citizenship record to earn a low grade. However, if the grades are supposed to be indicative only of academic achievement, such factors as attitude, absenteeism, tardiness, etc., should not be reflected. It is desirable for parents and others to be informed of such information; however, it should be reported separately. Gronlund[7] stressed that interpretation of single letter grades becomes hopelessly confused when factors other than achievement are included.

Teachers, students and parents within a department, school, or district can collectively identify standards or criteria in the form of qualitative statements that describe the level of achievement to be reached for each letter grade.[8] These standards may be applied in classes where students are grouped according to ability levels. It may be realistic to expect a higher number of students in the accelerated classes to achieve the standards set for higher grades than students in the slower classes. The same standards can also be applied in classes where mastery learning or contract learning systems are applied.

Relative Value of Scores

It is likely that over a period of time students may accumulate a massive number of scores, letters, minuses and pluses, or whatever symbols one chooses to record as evaluative data. There may be a record of scores earned by students on written exams and quizzes, performance tests, simulations, games, class discussion and direct observation techniques.

It is not likely that each of the tasks that students performed to earn these scores is of equal merit in terms of difficulty of tasks, relationship to objectives, or significance of content. Some distinctions can be made between the various scores to indicate the degree of importance of each task. Each score may be assigned a weight value based on its relative importance. Data reflecting student achievement on the most important objectives may be given greater weight than achievement on less important objectives.

Teachers, with the help of the students, can make decisions regarding the relative importance of instructional objectives prior to instruction. Students need to be informed of the relative importance of the instructional objectives since they will serve as an important guide in the evaluative process.

Concluding Comment

Evaluation in education is difficult and our measures are imperfect. The same is true for physicians as they try to judge our health status and diagnose our ills. Still they, and we, have to use what we have and do the best we can as we continue to improve our skills, our knowledge and our tools.

Perhaps one of our tasks as we make judgments is to be cognizant of the ways in which they can benefit the learner yet be ever mindful of the possible margin of error.

---

[7]Norman E. Gronlund, *Measurement and Evaluation in Teaching*, 2nd ed. (New York: The Macmillan Company, 1971), p. 488.

[8]Richard H. Lindemon, *Educational Measurement* (Glenview, Ill.: Scott, Foresman and Company, 1967), pp. 143-144.

PUBLICATIONS AVAILABLE

THE MASCULINE LOOK IN HOME ECONOMICS, B. Greenwood, J. Dowell, 1975
(Stock No. 261-08422).....$2.50
Boys tend to be oriented toward practical and competitive activities, basic skills and the sciences. This bulletin provides some answers to the question of how best a home economics teacher can adapt her teaching to challenge students of both sexes in the same class.

FAMILY LIFE EDUCATION: FOCUS ON STUDENT INVOLVEMENT, K. B. Green, 1975
(Stock No. 261-08420).....$4.00
Gives teachers and teacher educators help in turning listless students into stimulated committed learners. Step by step the teacher is led through the conceptual approach to teaching, fitting each of the components together for a role of mentor, stimulus, resource provider, counselor. Examples are given for each component. Bibliographic references for extra help are suggested.

NEW DIMENSIONS IN CREATIVITY, D. Gade, 1974 (Stock No. 261-08418).....$7.00
Provides a systematic approach for individuals and groups to develop their creative potential. Part one groups to develop the creative potential. It includes "Notes to the Teacher" and five teaching packets. Each packet includes lesson outlines, a narration for use as a script or as a reading assignment, and materials specified for use in class activities. Part two, for student use, includes five sections which correspond to the five teaching packets in part one.

CHOOSING TECHNIQUES FOR TEACHING AND LEARNING, H. T. Spitze, 1970
(Stock No. 261-08402).....$2.50
Examines the teaching-learning process and appropriate techniques for aiding learning. Techniques are categorized according to real life situations, simulation of reality, and abstractions from reality. Individualized instruction is a special topic.

TEACHING FOR CHANGED ATTITUDES AND VALUES, J. B. Ruud, 1972
(Stock No. 265-08378).....$2.50
A number of teaching techniques are discussed such as: role playing, surveys of attitudes, etc. Theories of several social psychologists are presented. Suggestions are made of how they be used as guidelines for teachers in teaching changed attitudes.

DETERMINING EFFECTIVENESS OF TEACHING HOME ECONOMICS, H. Chadderdon, 1971
(Stock No. 261-08408)..... .....$2.50
A discussion of evaluation techniques including steps in evaluation, purposes, improvement of evaluative practices, and pupil participation.

TEACHING FOR EMPLOYABILITY, P. D. Murphy, 1973 (Stock No. 261-08414).....$1.00
Concerns concepts in teaching for employability most widely accepted by educators as essential to the vocational curriculum. Topics covered are how to use these concepts in the classroom, various teaching techniques, techniques for attitude assessment, and how an individual teacher can evaluate his own program.

CLOTHING FOR YOUNG MEN, D. S. Lyle, 1970 (Stock No. 261-8302).....$2.50
Focuses on young men and their clothes. Fashion trends, selection and care of clothing are highlighted as well as information on new textiles and community resources available to the teacher.

EDUCATION FOR EFFECTIVE PARENTHOOD, K. S. Hughes, 1969 (Stock No. 261-08400).....$1.00
Suggests ways of initiating an effective parenthood education program with a laboratory for the young child as the "educational center" of that program. Describes successful programs in high schools that have established the value of the nursery laboratory in educating youth for their role as parents.

To order books indicate title, stock number, and number of copies desired. All orders must be prepaid except those on official order forms. Prices quoted are effective until January 1, 1977.

Mail to: Home Economics Education Association
1201 Sixteenth St., N.W.
Washington, D.C. 20036